

Kataloge kleinräumiger kommunalstatistischer Daten im Vergleich: was können KOSTAT, IRB, Urban Audit?

Gutfleisch, Ralf; Sturm, Gabriele

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Gutfleisch, R., & Sturm, G. (2013). Kataloge kleinräumiger kommunalstatistischer Daten im Vergleich: was können KOSTAT, IRB, Urban Audit? *Informationen zur Raumentwicklung*, 6, 471-491. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-58296-9>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:
<https://creativecommons.org/licenses/by-nc-nd/4.0>

Kataloge kleinräumiger kommunal- statistischer Daten im Vergleich – was können KOSTAT, IRB, Urban Audit?

Ralf Gutfleisch
Gabriele Sturm

1 Hintergründe kommunaler Datensammlungen

In aufgeklärten, demokratischen Gesellschaften begründen sich politische Entscheidungen, wo es möglich ist, auf empirischen Analysen erkundbarer Realitäten. Solche Analysen folgerten im 19. und in der ersten Hälfte des 20. Jahrhunderts häufig aus augenscheinlichen, eher journalistischen Berichten oder qualitativen Einzelfallstudien.

Zu den frühen empirischen Untersuchungen städtischer Räume zählen die Untersuchungen am Institut für Anthropologie und Soziologie der Universität von Chicago. Dort wurde 1925 erstmals der Versuch unternommen, städtische Teilgebiete anhand statistischen Datenmaterials zu differenzieren. Erst in den 70er Jahren ermöglichte die technische Weiterentwicklung, größere Datenmengen zu speichern und die bis dahin entstandenen ausgefeilten statistischen Analysemethoden anzuwenden. So werden für die Suche nach Generalisierungen (Induktion)¹ bereits seit dem 19. Jahrhundert Surveys eingesetzt. Auf Grundlage der heutigen Datenbanktechniken sind in Survey-Manier zunehmend Datensammlungen bzw. Kataloge² aus Umfragen, amtlichen Statistiken oder bestimmten Messpunkten etc. entstanden. Diese dienen einer kontinuierlichen und möglichst umfassenden Beobachtung gesellschaftlicher, natürlicher oder technischer Prozesse.

Das politische Umfeld war in jener Zeit u. a. durch Planungseuphorie gekennzeichnet: Ein Umbau der Gesellschaft erschien machbar und das Konzept der Lebensqualität trat in den Vordergrund planerischer Entwürfe. Entsprechend erlebten in den Wirtschafts- und Sozialwissenschaften Sozialindikatoren einen Boom. Sie galten als vergleichsweise bestes Mittel, um den Zustand und die Entwicklung einer Gesellschaft quantitativ zu ermitteln und mit anderen Gesellschaften zu vergleichen. Das Zusammenspiel von

gesellschaftlich weit verbreiteten Überzeugungen, sozialwissenschaftlichen Operationalisierungen und den neu entstehenden Möglichkeiten der Informationstechnik bildet bis heute ein Gerüst auch für Stadt- und Raumbeobachtung. Diese widmet sich der Beobachtung regionaler Lebensbedingungen und deren Veränderung mittels raumbezogener Indikatoren. Dazu bedarf sie zusätzlich zu den amtlichen/kommunalen oder privatwirtschaftlichen Daten – auch Geofachdaten genannt – der Festlegung eines räumlichen Bezugssystems – der Geobasisdaten.

Alle Daten mit Raumbezug können als Geodaten bezeichnet werden; sie sind mit einer bestimmten Lage auf der Erdoberfläche versehen. Geodaten gliedern sich in die Geobasisdaten, die in der Regel von den Vermessungsverwaltungen der Länder oder der Kommunen bereitgestellt werden, und die Geofachdaten, die aus unterschiedlichen Fachstatistiken gespeist werden, wie z. B. demografischen Daten, der Wahlstatistik oder einem Liegenschaftskataster. Dieser Sprachgebrauch ist stark durch Geoinformatik bzw. den Umgang mit Geodateninformationssystemen (GIS) geprägt und für Vertreterinnen und Vertreter anderer wissenschaftlicher Disziplinen ungewohnt bis unbekannt. Wir verwenden die Begrifflichkeit hier, um die grundsätzliche Bezogenheit raumbezogener Merkmale und ihrer Lokalisierung zu betonen.

Die Wahl des Raumbezugs ist von entscheidender Bedeutung dafür, wie die für die regionale Entwicklungspolitik verfügbaren Mittel (z. B. der Wirtschaftsförderung, Verkehrserschließung, Agrarstrukturverbesserung usw.) auf die einzelnen Teile des Staatsgebiets verteilt werden. Wichtige Arbeitsmittel für die Raumbeobachtung sind deshalb geografische Informationssysteme und die Daten der amtlichen Statistik, die in Indikatoren einfließen.

Dr. Ralf Gutfleisch
Bürgeramt, Statistik
und Wahlen
Zeil 3
60313 Frankfurt
E-Mail: ralf.gutfleisch@
stadt-frankfurt.de

Dr. Gabriele Sturm
Bundesinstitut für Bau-, Stadt-
und Raumforschung (BBSR)
im Bundeamt für Bauwesen
und Raumordnung
Deichmanns Aue 31–37
53179 Bonn
E-Mail: gabriele.sturm@
bbr.bund.de

(1) Strukturlogisch werden drei verschiedenen Schlussweisen unterschieden: Abduktion = Zusammenhangsaussagen anhand eines Falls; Deduktion = theoretische Aussagen; am empirischen Befund überprüft; Induktion = generalisierende Aussagen/Regelhaftigkeiten anhand zahlreicher Fälle (siehe auch Sturm 2005).

(2) Die empirische Sozialforschung versteht unter Survey eine Erhebung, die möglichst alle oder eine repräsentative Auswahl der Merkmalsträger innerhalb eines festgelegten Bereiches erfasst. Im Bereich der Geowissenschaften verweist das Surveying auf ein Vermessungsnetz. In jedem Fall ist ein Survey auf einen allgemeinen Überblick ausgerichtet. Die Datenbanken, in denen solche Überblicksdaten abgelegt sind, werden in diversen Disziplinen Katalog genannt. Diesen Sprachgebrauch übernehmen wir hier für eine Sammlung mit unterschiedlichen aufeinander bezogenen Daten insgesamt. Geht es hingegen um bestimmte Referenzdaten oder Merkmalslisten, die einem solchen Katalog zuzuordnen sind, bleiben wir bei den differenzierteren Bezeichnungen.

(3) Repräsentation bedeutet ganz allgemein, dass etwas durch etwas anderes vertreten oder dargestellt wird: Aussagen in Form von Indikatoren oder empirischen Befunden – soweit diese wahr sind – repräsentieren etwas, das auch unabhängig von diesen Aussagen existiert, z. B. Gegenstände, Sachverhalte, Beziehungen etc. Zugleich ist allerdings zu bedenken, dass die Darstellung bzw. die Aussage das Repräsentierte verändert oder beeinflusst, es i. d. R. keine Repräsentation ohne Rückwirkung auf das Repräsentierte gibt. In den Kommunikationswissenschaften besagt das Repräsentationsmodell, „dass bestimmte Texte ein verlässliches Abbild der dahinterstehenden Realität abgeben und bei Gültigkeit der Annahme, dass keine bewussten Verzerrungen im Text enthalten sind, ein Schluss von Text auf Kontext gültig ist.“ (Atteslander 1995: 241).

Auf Bundesebene werden Daten auf unterschiedlichen Aggregatebenen gesammelt – neben einem umfangreichen Datensatz auf Kreisebene liegen auch zahlreiche Daten auf Ebene der Städte und Gemeinden vor. Diese Daten beziehen sich in der Regel auf die Städte als Ganzes, also auf das administrative Gebiet in ihren Verwaltungsgrenzen. Zudem führen die Kommunen selbst im Rahmen ihrer Kommunalstatistik für eigene Zwecke je nach Ressourcen mehr oder weniger umfangreiche kleinräumige Sammlungen durch. Mit Hilfe geeigneter Indikatoren soll auf allen Ebenen ermöglicht werden, entstehende Probleme frühzeitig zu benennen oder bereits entstandene Missstände aufzuzeigen, um auf dieser Grundlage dann mögliche planungspolitische Interventionen zu entwickeln.

Die empirischen Befunde bzw. Indikatoren und Informationen sind abhängig von der Bearbeitung und der Qualität der zu Grunde gelegten Daten. Und dafür ist zuerst zu fragen, was die erhobenen Daten repräsentieren³. So repräsentieren Umfragedaten, was die Befragten denken, meinen, urteilen und/oder einschätzen, und kaum, was sie tatsächlich tun oder getan haben. Kommunale und amtliche Statistiken repräsentieren, was in den verschiedenen Verwaltungen gesammelt und gezählt wird bzw. werden kann. Dafür erscheint es sinnvoll, kurz in die Vergangenheit zu schauen, denn die Wurzeln amtlicher Statistik reichen weit zurück:

- Volkszählungen in den Jahrhunderten vor unserer Zeitrechnung (Babylon, Ägypten, China, römisches Reich) dienten z. B. mittels der Aufstellung eines Vermögenskatasters der Steuer- und Finanzverwaltung. Weiterhin wurden sie für die Planung des Militärdienstes eingesetzt.
- Einfuhrlisten von zollpflichtigen Waren, Vorratslisten oder Handelsstatistiken dienten nicht nur der Finanzverwaltung oder der Regelung von Bevölkerungszügen, sondern führten nach und nach auch zur Erfassung der Infrastrukturausstattung: Im 18. Jahrhundert wurden z. B. in Preußen Daten zu Hausbestand und Kommunalfinanzen, Viehbestand, Aussaat, Getreidepreisen, Flachs- und Tabakanbau, Fabriken, Hütten- und Bergwerken, Schifffahrt und Handel erhoben.

- Von einer amtlichen, systematischen Statistik mit aufeinander abgestimmten Erhebungen kann man in Deutschland seit Gründung der zentralen deutschen Statistik und ihrer Institutionen im 19. Jahrhundert sprechen – also des Kaiserlichen Statistischen Amtes, auf Ebene der Einzelstaaten der statistischen Landesämter und auf Ebene der Kommunen der unabhängigen statistischen Büros. Sie legten den Grundstein für eine analytische Berichterstattung zu unterschiedlichen Themenbereichen wie sie bis heute in den Veröffentlichungen der statistischen Ämter vollzogen wird (vgl. u. a. Gerß 2010 oder Sturm 2012b).

Statistische Sammlungen sind auch heute noch hinsichtlich Merkmalsauswahl und Festlegung der Merkmalsausprägungen von diesen Ursprüngen geprägt. Aktuell werden ihr Umfang und ihre Struktur zudem von politischen, ökonomischen und planerischen Interessen beeinflusst, die sich je nach Thema oder nach Ebene in der föderalen Struktur unterscheiden können. All dies gilt folglich auch für Kataloge mit Kommunaldaten. Während aufgrund gesetzlicher Vorlagen die Bundes- und Landesebene – wenn auch zeitlich verzögert – stets abgestimmte Kataloge bis auf Kreisebene bereitstellen müssen, fehlt ein abgestimmter Pflichtkatalog zwischen den Städten – insbesondere was kleinräumige Daten betrifft. Dies ist auf das dezentrale Städtesystem mit variierenden Stadtcharakteristika und auf die unterschiedlichen technischen und personellen Ressourcen der Städte zurückzuführen. Schließlich lässt auch die kommunale Selbstverwaltung der Städte einen Pflichtkatalog nicht zu. Um dennoch Stadtvergleiche anstellen zu können, entstanden auf freiwilliger Basis diverse Kataloge mit kommunalen Geodaten.

Die drei derzeit gepflegten Kataloge mit kleinräumigen Kommunaldaten sollen hier vorgestellt und ihre Potenziale einem breiteren Wissenschaftspublikum vorgestellt werden. Dazu werfen wir zunächst einen Blick in die Entstehungsgeschichte der Kataloge und stellen ihre aktuelle Gestalt vor. In weiteren Kapiteln diskutieren wir Datenqualität und Synergiepotenziale. Der Beitrag endet mit Überlegungen zur künftigen Gestaltung der Kataloge.

2 Entstehungszusammenhänge kommunaler Sammlungen kleinräumiger Daten

2.1 Rahmenbedingungen

Kataloge mit Stadt- und Stadtteildaten sind in ihrer Vielfalt und ihren Unterschieden nur aus dem Prozess ihrer Entwicklung heraus zu verstehen:

- Dem Zeitgeist entsprechend wurde in der damaligen Bundforschungsanstalt für Landeskunde und Raumordnung (BfLR) Mitte der 70er Jahre die Laufende Raumbeobachtung (LRB) als Informationssystem für die raumorientierte Forschung und Politikberatung auf- und ausgebaut. Die Führung eines „Informationssystems zur räumlichen Entwicklung im Bundesgebiet“ wurde später bei der Fusion der BfLR mit der Bundesbauverwaltung zum Bundesamt für Bauwesen und Raumordnung (BBR) im Raumordnungsgesetz (ROG vom 18.08.1997, §18 (5)) als Daueraufgabe festgelegt. Für dieses räumliche Informationssystem wurden zunächst nur flächendeckende Daten z. B. auf Gemeinde- oder Kreisebene zusammengetragen.
- Zu Beginn der 80er Jahre fand ein Politikwechsel statt: Raumordnungspolitik wurde nicht länger als integrierte staatliche Entwicklungsplanung gesehen, sondern Raumordnungspolitik sollte eher problemorientiert als „Krisenmanagement“ eingesetzt werden. Zugleich gewannen Probleme der Agglomerations- und Stadtentwicklung an Bedeutung und verlangten nach verlässlichen Informationen. Für die neuen städtebaulichen Aufgaben fehlte jedoch ein Informationssystem. In der Folge wurde die LRB der BfLR um das erste Kooperationsprojekt „Innerstädtische Raumbeobachtung“ (IRB) 1985/86 ergänzt und so zur Laufenden Raum- und Stadtbeobachtung. Das Ziel der darauf fußenden Politikberatung lautete: Koordination durch Information. Für ein raumbezogenes Informationssystem verbreitete sich die Idealvorstellung, Indikatoren von der Ebene der Adresse bis zur Ebene des Nationalstaates (bzw. der EU) durchgängig zur Verfügung zu haben.
- Seit den 70er Jahren hat die Entwicklung der Informationstechnologie den Aufbau großer Datenbanken und die An-

wendung komplexer statistischer Analysen immer leichter werden lassen. Nur aufgrund dieser Rahmenbedingung ließen sich die derzeit entstehenden Kataloge mit raumbezogenen Daten realisieren. In den 80er Jahren fand dann der Wechsel von einer Großrechnerdatenhaltung auf eine auf zahlreiche Einzelarbeitsplätze verteilte Datenhaltung statt. In der Folge entstanden – je nach fachlichen Anforderungen und persönlichen Hintergründen unterschiedlicher Benutzer – verschiedenen Formate der Datenhaltung und -bearbeitung. Inzwischen erwachsen aus dieser dezentralen Datenhaltung zunehmend Probleme, wenn es um Austausch, Zusammenführung oder Vernetzung von Datenbeständen oder Analysen geht. Solches gilt vielerorts bereits innerhalb einer Institution und nicht erst bei der Zusammenarbeit verschiedener Datenhalter. Nicht nur bedarf es zeitaufwändiger Übertragungs- und Anpassungsarbeiten inklusive der entsprechenden Fehlerquellen, sondern zahlreiche Daten und Informationen werden zudem parallel gesammelt und aufbereitet, u. a. aufgrund mangelnder Kenntnis über deren Existenz oder unzureichender Information über ihre Struktur und Verfügbarkeit. Um diese Probleme mittelfristig zu lösen, wird aktuell an einer Standardisierung von Datenmodellen⁴ wie von Schnittstellen und Zugangswegen, z. B. in der Form von Webdiensten, und einer systematischen Erfassung von Metadaten gearbeitet. Auf europäischer Ebene dient diesem Anliegen u. a. die Richtlinie Infrastructure for SPatial InfoRmation in Europe (INSPIRE). Die Umsetzung dieser Richtlinie erfolgt in Deutschland durch Geodateninfrastruktur Deutschland (GDI-DE).

- Seit den 90er Jahren ist auf Seiten der Marktforschung ein ausgeprägtes Interesse an räumlich referenzierten Daten entstanden. Dabei geht die Entwicklung von Navigationsprogrammen für die Verkehrsführung Hand in Hand mit Standortanalysen z. B. für die Ansiedlung von Filialbetrieben oder für Vermarktungsstrategien. Parallel zur Kommerzialisierung räumlicher Informationen wurde in zahlreichen Behörden die Haushaltsführung umgestellt und viele der als weniger zentral geltenden Tätigkeitsbereiche wurden privatisiert. Dies führte einerseits bei der nun notwendigen Vergabe von Aufträgen seitens der Ämter zu einem bislang nicht üblichen Ausloten von

(4)

Ein Datenmodell bzw. Datenbankmodell ist die theoretische Grundlage für ein Datenbanksystem und bestimmt, auf welche Art und Weise Daten in einem Datenbanksystem gespeichert und bearbeitet werden können. Es legt die Infrastruktur fest, die ein bestimmtes Datenbanksystem anbietet. Am weitesten verbreitet ist derzeit das relationale Datenbankmodell.

Angebot und Nachfrage und andererseits auch bei öffentlichen Einrichtungen zur Notwendigkeit, die eigenen Produkte vermarkten zu müssen. Damit mussten und müssen sich auch öffentlich finanzierte Kataloge hinsichtlich ihres Nutzens hinterfragen lassen.

- Zugleich wächst in Politik und Verwaltung das Interesse an kleinräumigen Daten, um adäquate Entscheidungen treffen zu können. Die Anwendungsgebiete reichen hier von Sozialraumanalysen über Einzelhandelskonzepte bis hin zu Wahlanalysen. Häufig reichen dabei die Daten auf Ebene der administrativen Gebietsgliederung nicht mehr aus, um die Sachverhalte eindeutig abzubilden. Die Verwaltung muss immer häufiger auf die Adressebene zurückgreifen und zu analytischen Zwecken neue Gebietsabgrenzungen generieren.

Ein Blick über die vergangenen 40 Jahre zeigt, dass während dieser Zeit die Interessen an georeferenzierten Daten wie deren Nutzerinnen und Nutzer in Politik und Wirtschaft andere geworden sind.

Zudem werden im Prozess des gesamtgesellschaftlichen, demografischen Wandels immer wieder andere Themen wichtig, obwohl ihre Wahrnehmung und Erfassung oft noch alten, nicht unbedingt problemadäquaten Mustern folgen. So unterscheiden viele Analysen auch mehr als 20 Jahre nach der deutschen Einheit noch nach Ost- und Westdeutschland entlang der ehemals innerdeutschen Grenze, erneuern damit altes Ost-West-Denken und verdecken andere räumliche Unterschiede. Oder vor dem Hintergrund, dass Deutschland sich als Einwanderungsgesellschaft erwiesen hat, war und ist von verschiedenen Akteurinnen und Akteuren ein sehr unterschiedlicher Sprachgebrauch festzustellen. In den meisten Statistiken und den daraus abgeleiteten Analysen wird immer noch Staatsangehörigkeit – also der Indikator „Ausländer“ – verwendet. Vor dem Hintergrund eines Zuwachses an Spätaussiedlerinnen und -aussiedlern in den 90er Jahren, der Etablierung der zweiten und dritten Generation der ehemaligen „Gastarbeiter“ sowie der steigenden Zahl von Einbürgerungen erscheint dies schon lange nicht mehr adäquat. Auf Grundlage des Mikrozensus hat inzwischen die Bundesstatistik zwar als al-

ternativen Indikator den Migrationshintergrund diskursfähig gemacht, zwischen den Institutionen existiert dafür jedoch noch keine einheitliche Definition.

In den Städten, in denen Zuwanderung eine wichtige Rolle in der Bevölkerungsdynamik spielt, setzen sich die Verantwortlichen schon seit 20 Jahren mit der Definitionsproblematik auseinander. Verstärkt wurde die Diskussion um die Abgrenzung der Migranten durch die zum 1. Januar 2000 in Kraft getretene Novellierung des Staatsangehörigkeitsrechts. Aus diesem Grund wurden verschiedene Verfahren zur Datenbereitstellung des Migrationshintergrundes entwickelt (u. a. das Programm MigraPro). Aber auch in anderen Bereichen in denen demografische Indikatoren eine wichtige Rolle spielen, wie z. B. in der Schul- oder Arbeitslosenstatistik, entstand zunehmend der Druck, neue Abgrenzungen zu finden. Dies führte schließlich dazu, dass je nach Bedürfnislagen und Kontext der Migrationshintergrund unterschiedlich definiert und operationalisiert wurde und wird. Vergleiche zwischen den unterschiedlichen Erhebungen und Berechnungen sind quasi unmöglich (vgl. VDst 2013)

Weiterhin ist bei den hier vorzustellenden Katalogen die Zeitreihenproblematik zu berücksichtigen. Die Sammlungen stellen auch ein Archiv der gesellschaftlichen Veränderungen dar. So wäre ein Austausch bzw. eine Neudefinition einer Variablen mit einem Einschnitt verbunden und würde einen Bruch in der Beobachtung über die Zeit bedeuten. Rückrechnungen und Umschätzungen für die Vergangenheit können nur in Ausnahmefällen erfolgen. Umstellung oder Neudefinitionen werden daher nur sehr selten vorgenommen. Indikatoren auf Basis der eingestellten Daten erscheinen deshalb manchmal der gesellschaftlichen Wirklichkeit hinterherzuhinken.

All dies ist für die im Weiteren vorzustellenden Kataloge mitzubedenken.

2.2 Die Innerstädtische Raumbeobachtung – IRB

Die Innerstädtische Raumbeobachtung (IRB) bildet das Pendant zur Laufenden Raumbeobachtung (LRB) des BBSR – als Grundlage für eine überörtlich vergleichende Stadt- und Raumforschung. Ähnlich wie im Rahmen der LRB bundesweit die flächendeckenden regionalstatistischen Daten für Kreise und Gemeinden zusammengestellt werden, versammelt die IRB städtestatistische Aggregatdaten für vergleichbare Raumeinheiten unterhalb der Gemeindeebene. In der Zusammenschau bietet sich so die Möglichkeit, speziell auch über Stand und Entwicklung der deutschen Großstadtregionen berichten und Zusammenhänge analysieren zu können.

Start:

Die IRB wurde im März 1986 als Kooperationsprojekt des Deutschen Städtetags (DST), des Verbands Deutscher Städtestatistiker (VDSt), des KOSIS-Verbunds der Kommunen⁵, des Deutschen Instituts für Urbanistik (DifU) und der BfLR gegründet. Grundlage bildete eine Rahmenvereinbarung über die Kooperation bei Aufbau und Nutzung einer überörtlichen städtestatistischen Datenbasis für vergleichbare Raumeinheiten unterhalb der Gemeindeebene und bei der Nutzung der LRB.

Verlauf:

Anfangs gehörten 27 Städte der Kooperation an, die zusammen mit den die Stadtforschung repräsentierenden Vertreterinnen

(5)
KOSIS ist die Abkürzung für „Kommunales Statistisches Informationssystem“ – vgl. www.staedte-statistik.de/KOSIS.html.

Karte 1
IRB-Städte in den BBSR-Großstadtregionen

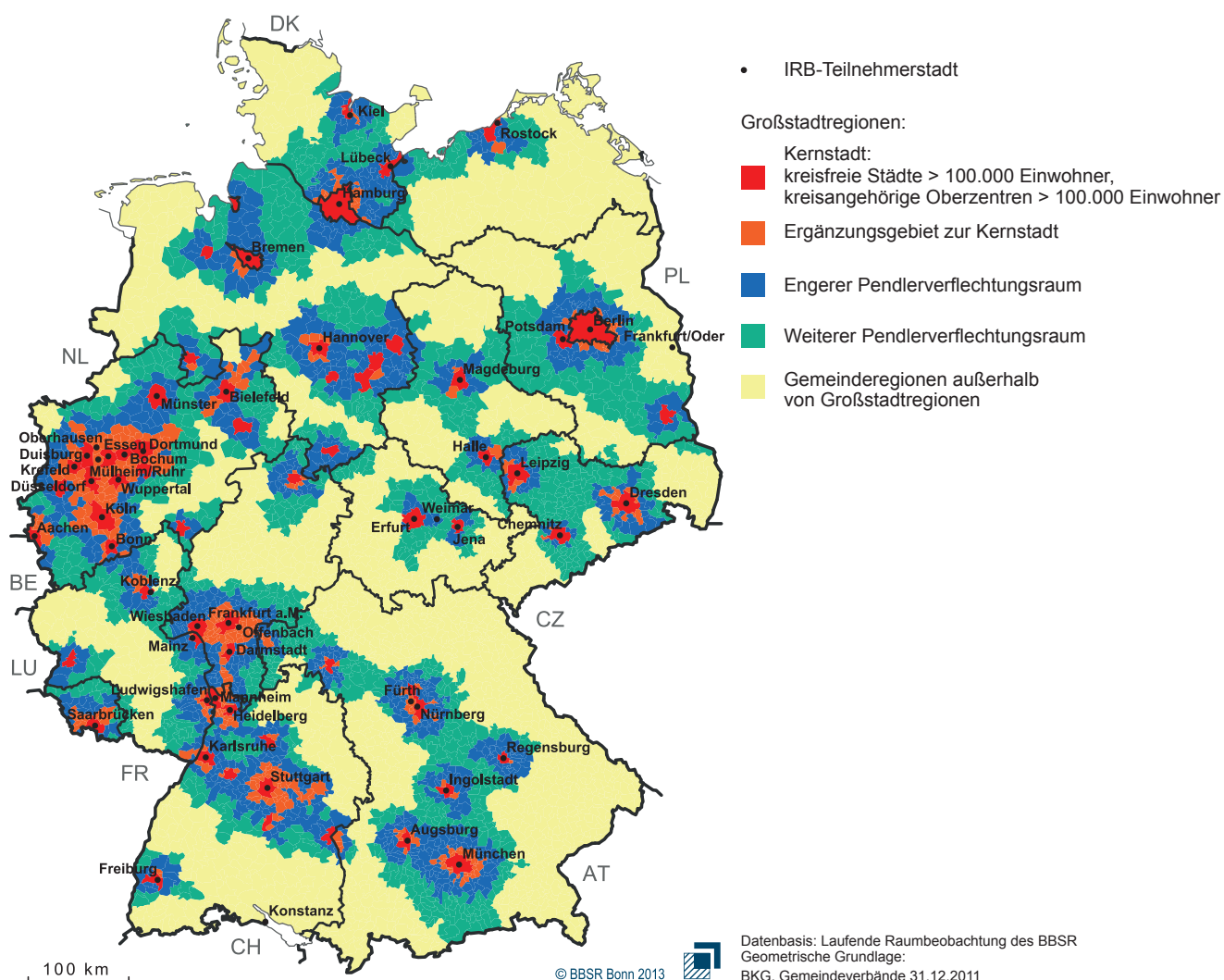
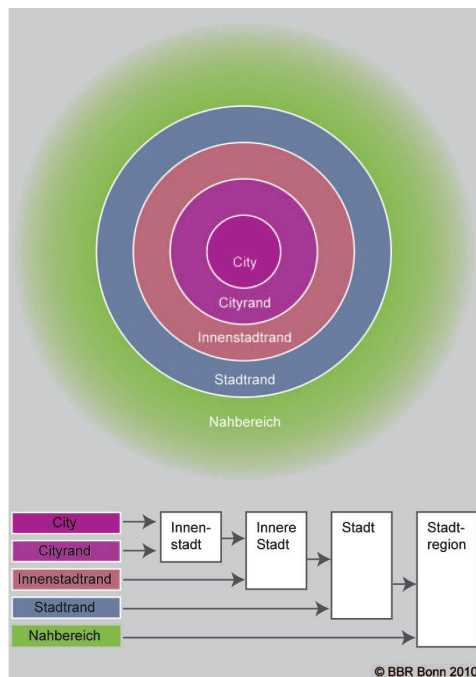


Abbildung 1
IRB-Lagetypen gemäß klassischer Zentren-
konzeption



und Vertretern ein gemeinsames Merkmalprogramm aushandeln. Bis 1993 waren dem Projekt 35 Teilnehmerstädte beigetreten. Anlässlich der Datenlieferung für das Jahr 1998, die die Mitarbeiter des damals neu fusionierten BBR organisierten, wurde die von den Städten zunehmend als unprofitabel angesehene IRB zugunsten des neu gegründeten KOSTAT-Katalogs abgebrochen. Da dessen Datenspektrum für Stadtforschungsbedarfe jedoch auf Dauer als unzureichend empfunden wurde, startete mit dem Erhebungsjahr 2002 die Neugründung bzw. Wiederbelebung der IRB.

Organisation:

Geführt wird dieser Katalog derzeit im Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR). Erfahrungsaustausch, Absprachen und Weiterentwicklung des Projekts finden bei jährlichen Mitgliederversammlungen und in den Zwischenzeiten per E-Mail-Netzwerk statt. Darüber hinaus findet ein Austausch über Entwicklungsmöglichkeiten und eine Abstimmung mit anderen Katalogen in enger Zusammenarbeit mit dem VDSt und dem KOSIS-Verband statt. Alle Projektbeteiligten dürfen den Katalog für ihre Analysen nutzen. Eine wissenschaftliche Nutzung durch Dritte ist stark reglementiert.

aktuell:

Der Katalog enthält ausschließlich Geofachdaten. Seit der Datenlieferung für das Jahr 2011 nehmen 51 Städte mit insgesamt 2 889 bewohnten Raumeinheiten (statistischen Bezirken, Stadtteilen, Ortsteilen)⁶ an der neuen IRB teil. Sie beliefern jährlich 30 Tabellen bzw. Aggregatmerkmale mit insgesamt mehr als 400 Merkmalsausprägungen. Für Analysezwecke sind die Raumeinheiten als Merkmalsträger einem innerstädtischen Lagetyp zugeordnet, der sich dem Boustedtschen Zentralitätskonzept entlehnt: Unterschieden wird nach City, Cityrand, Innenstadtrand und Stadtrand. Die Kategorisierung der Raumeinheiten in Lagetypen wurde in jeder Stadt durch Zusammenwirken der Stadtplanungs- bzw. Stadtentwicklungsämter mit den statistischen Ämtern vorgenommen. Sie entsprechen somit den je lokalen Planungskriterien. Mit den Lagetypen steht eine weitere Aggregatebene zur Verfügung, die sich – trotz der sehr einfachen Unterscheidung – als analytische Differenzierung bewährt hat.

2.3 Die Kataloge im KOSIS-Verbund

Der KOSIS-Verbund ist eine kommunale Selbsthilfeorganisation, die mit Unterstützung des Deutschen Städtetags Kooperations- und Gemeinschaftsprojekte organisiert. Diese Projekte sind darauf gerichtet, die DV-Instrumente vor allem für kommunale Statistik, Stadtforschung und Planung sowie Wahlen kooperativ und damit kostensparend zu organisieren, d. h. zu entwickeln oder zu beschaffen, zu warten und zu pflegen. Der KOSIS-Verbund hat es auch übernommen Daten der kommunalen Statistik – insbesondere die kleinräumigen Gliederungssysteme und kleinräumig gegliederte Daten – überörtlichen Interessierten zugänglich zu machen. Insgesamt fördert er mit seinen Aktivitäten die inhaltliche und technische Standardisierung und unterstützt das kommunale Informationsmanagement und mit ihm die Leistungsfähigkeit der Kommunalverwaltung. Träger des KOSIS-Verbundes ist der Verband Deutscher Städtestatistiker (VDSt).

Der VDSt ist ein eingetragener Verein, der sich zum Ziel gesetzt hat, die Städtestatistik und Stadtforschung in der Kommunalverwaltung, in Wissenschaft und Praxis zu fördern. Zudem beobachtet er die

(6) Die 2 889 bewohnten Raumeinheiten / Stadtteile der IRB haben 2011 zwischen 3 und 86 902 Einwohner; bei 10 bis 447 Raumeinheiten pro Stadt kommen im Durchschnitt etwa 7 150 EW auf jede Raumeinheit.

Entwicklung der Arbeitsbedingungen und Rechtsgrundlagen der Kommunalstatistik und unterstützt den Deutschen Städtetag (DST) bei seiner Lobbyarbeit für die Städte. Arbeitsformen des VDSt sind regionale und fachliche Arbeitsgemeinschaften sowie regelmäßige Fachtagungen und Publikationen.

Außer den beiden Daten sammelnden KOSIS-Gemeinschaften AG Kostat und Urban Audit, die hier näher vorgestellt werden, bestehen weitere selbstständige Gemeinschaften mit unterschiedlichen Schwerpunktthemen und entsprechenden Programmen:

- DUVA entwickelt Lösungen zum Informationsmanagement,
- HHStat zur koordinierte Haushalte- und Bevölkerungsstatistik,
- SIKURS zur regionalisierten Bevölkerungsprognose,
- AGK zu Adresszentraldatei, Gebäudedatei und Kleinräumiger Gliederung,
- KORIS zu Kommunalen Raumbezügen und zur Raumanalyse,
- KOWAHL zur Wahlorganisation, Ergebnisanalyse und -präsentation.

2.3.1 AG KOSTAT

Die Arbeitsgemeinschaft Kommunalstatistik (AG KOSTAT) betreibt die Sammlung von kommunalstatistischen, insbesondere kleinräumig gegliederten Daten und stellt sie für kommunale Analysen, aber auch für andere öffentliche und private Auswertungszwecke gegen Nutzungsentgelt zur Verfügung.

Start:

Aus den beiden bis 1999 durchgeführten KOSIS-Projekten Innerstädtische Raumbeobachtung (IRB) und Weitergabe kleinräumiger Gliederungssysteme (WKG), beide mit Sitz in Stuttgart, ging zunächst die KOSTAT-DST GmbH hervor. Die Projekte IRB und WKG waren in ihrem Verlauf den Verantwortlichen als immer schwieriger umsetzbar und den Städten als Datenlieferanten als immer größere Belastung erschienen. Dies ist vor allem unter dem Druck der Kosteneffizienz und Vermarktung kommunaler Produkte in den Städten zu sehen. Die neu gegründete KOSTAT-DST GmbH hatte daher zum Ziel, kommunalstatistische Daten

und kleinräumigen Gliederungssysteme an unterschiedliche Zielgruppen zu verkaufen und sich damit auf dem sich bildenden Geodaten-Markt zu platzieren. Erwartet wurde ein fester Abnehmerkreis, der weiter zunehmen würde, wenn nur die Daten aktuell, vollständig und für den überwiegenden Teil der Großstädte erhältlich wären. An der GmbH waren beteiligt der Deutsche Städtetag, die Städte Dortmund, Frankfurt a.M., Hannover, Leipzig, Lübeck, Oberhausen und Nürnberg. Die Geschäftsführung fand im Fachbereich Statistik und Wahlen in Dortmund ihren Sitz.

Verlauf:

Um einen möglichst hohen Beteiligungsgrad unter den Städten zu erlangen, wurde als niederschwellige Einstiegsvoraussetzung der Datensatz auf wenige nachgefragte Daten (elf demografische Merkmale) begrenzt. Bereits 2002 musste jedoch festgestellt werden, dass die Umsatzerwartungen nicht erfüllt wurden. Um die erfolgreiche Aufbauarbeit und den gewonnenen Kundenkreis nicht zu gefährden, beschloss man nach der Liquidation der Gesellschaft, das Projekt mit geringeren Fixkosten in einer KOSIS-Gemeinschaft fortzuführen. So kam es zur Gründung der Arbeitsgemeinschaft Kommunalstatistik (AG KOSTAT) durch die Städte Bielefeld, Dortmund, Frankfurt a.M., Hannover, Nürnberg und Stuttgart unter Beteiligung des BBR.

Organisation:

Die Geschäftsstelle wechselte nach Frankfurt a. M., wo sie bis heute im Bürgeramt, Statistik und Wahlen ihren Sitz hat. Einmal im Jahr findet hier auch die Mitgliederversammlung statt. Das Ziel der AG ist es nach wie vor, kleinräumige kommunalstatistische Daten zu sammeln und einer interessierten Öffentlichkeit anzubieten. Aufgrund der zeitlich unterschiedlichen Aufbereitung der Melderegister werden die Datenlieferungen der KOSTAT-Städte bis Mitte des Folgejahres in der Geschäftsstelle gesammelt und dann im Herbst den Abnehmerinnen und Abnehmern aufbereitet zur Verfügung gestellt.

Für kommunale Analysen ist der Zugang zum KOSTAT-Katalog für die Lieferstädte kostenlos. Für öffentliche und private Auswertungszwecke werden die Daten gegen Entgelt zur Verfügung gestellt. Zu den Nut-

zern aus dem wissenschaftlichen Bereich zählen neben dem BBSR, das die Daten u. a. zur Kontrolle und Vervollständigung der IRB-Daten verwendet, universitätsnahe Forschungseinrichtungen. Die größere Abnehmergruppe stellen jedoch die privaten Markt- und Meinungsforschungsinstitute oder Firmen in deren unmittelbarem Umfeld dar, die in den letzten Jahren die KOSTAT-Daten mit wachsendem Zuspruch nutzen.

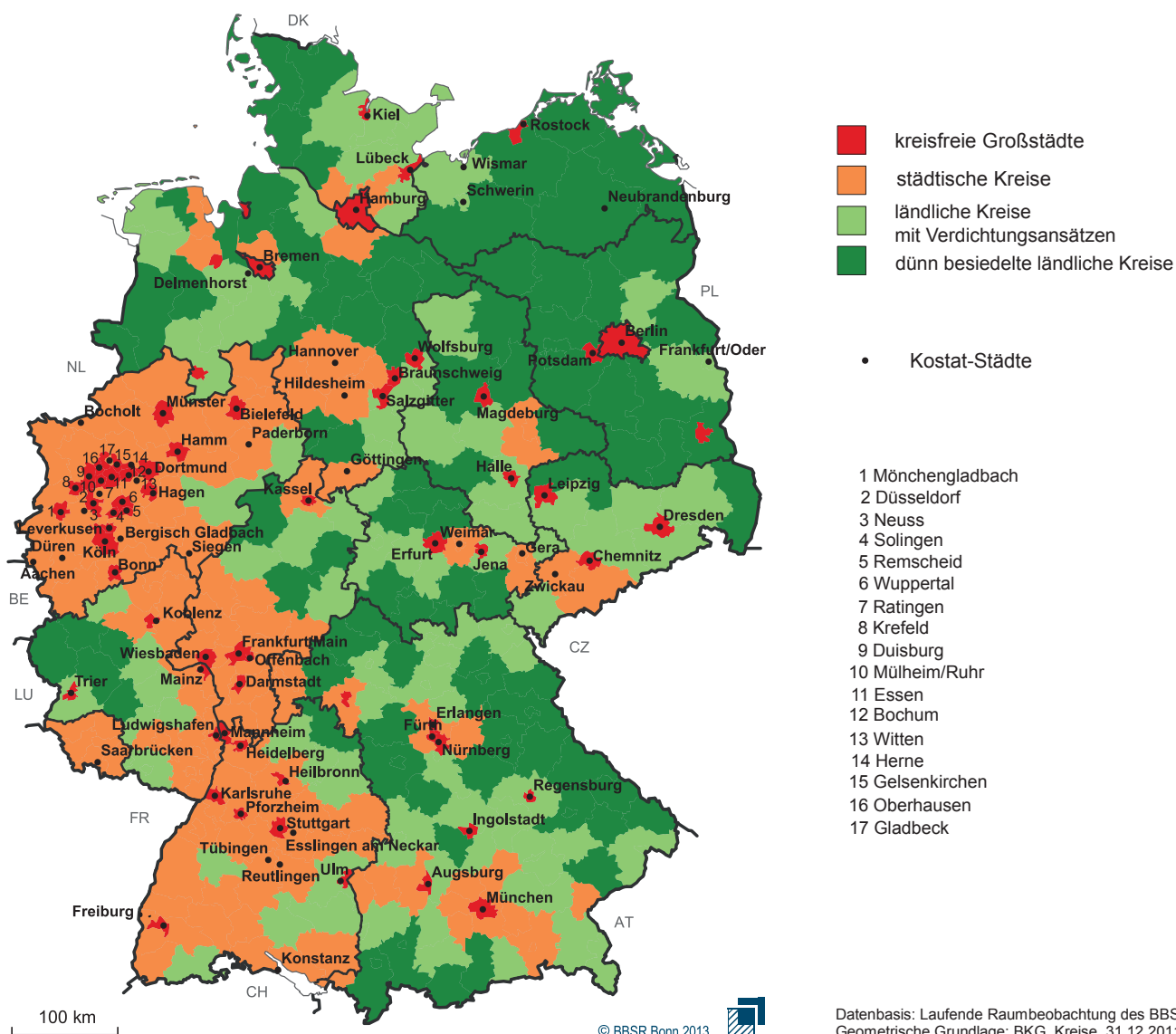
aktuell:

Etwa 100 deutsche Kommunen, darunter nahezu alle Großstädte beteiligen sich an dem Projekt. Die Großstädte mit mehr als 500 000 Einwohnern sind alle, die mit mehr als 100 000 Einwohnern fast vollständig ver-

treten. Von den großen Mittelstädten mit mehr als 50 000 Einwohnern beteiligt sich ein Fünftel an der Datensammlung. Die Einwohnerzahl aller KOSTAT-Städte entspricht etwa einem Drittel der Bevölkerung Deutschlands.

Der Katalog enthält sowohl Geofach- als auch Geobasisdaten. Die Geobasisdaten werden in Form von Straßenverzeichnissen und Geometrien geliefert. Jährlich werden von etwa 100 Lieferstädten Sachdaten überwiegend auf der dritten Gliederungsebene (kleinräumigste administrative Ebene: etwa 10 500 Raumeinheiten mit durchschnittlich etwa 3 000 Einwohnern) in den Katalog eingestellt. Dabei handelt es sich um drei

Karte 2
KOSTAT-Städte vor siedlungsstrukturellem Kreistyp des BBSR



Aggregatmerkmale mit insgesamt elf Merkmalsausprägungen, die von den Statistikstellen der beteiligten Städte zur Verfügung gestellt werden, nachdem sie aus dem Melderegister abgezogen wurden.

2.3.2 Urban Audit

Urban Audit ist ein Informationssystem zur Unterstützung des europäischen Städtevergleichs. Es verbindet eine Vielzahl von Datenlieferanten, Nutzern und Informationen mit dem Zweck, die Lebensqualität in den Städten der Europäischen Union (EU) beobachten, beurteilen und vergleichen zu können. Zudem soll der europäische Städtevergleich die Regionalpolitik der EU fundieren und begleiten sowie die nationale, regionale und lokale Städtepolitik unterstützen. Mit der Bereitstellung vergleichbarer Daten und Indikatoren wird also auch das Ziel verfolgt, Interventionsstrategien zur Minderung der erkannten Disparitäten zwischen und innerhalb der Städte zu entwickeln.

Start:

Bereits in den Jahren 1997 bis 2000 wurde im Auftrag der EU-Kommission und betreut vom Statistische Amt der Europäischen Union (EuroStat) sowie der Generaldirektion Regionalpolitik und Stadtentwicklung der Europäischen Kommission (GD Regio) ein Pilotprojekt unter dem Namen Urban Audit (UA) durchgeführt. Die damals gesammelten Daten bezogen sich auf das Jahr 1999. Auftragnehmer der Kommission war ein belgisch-britisches Konsortium von Instituten, die mit Hilfe sogenannter nationaler Korrespondenten (für die Bundesrepublik DIfU und KOSIS) das Pilotprojekt in den Mitgliedsstaaten durchführten. Insgesamt 58 europäische, darunter neun deutsche Großstädte waren damals an diesem ersten europäischen Städtevergleich beteiligt.

Verlauf:

2002 wurde eine Fortschreibung des UA-Projekts mit weiteren Städten bzw. städtischen Räumen, einem überarbeiteten Merkmalsprogramm sowie neuer Organisation der Durchführung im System der jeweiligen nationalstaatlichen amtlichen Statistik beschlossen. So wurde z. B. der Kreis der beteiligten Großstädte um Städte zwischen 50 000 und 250 000 Einwohnern erweitert. Am Urban Audit II (Bezugsjahr 2001) betei-

ligten sich 189 große und mittlere Städte der Mitgliedsstaaten der damaligen EU und 69 Städte der Beitrittsländer (gesamt: 258 Städte in der EU 27). 123 dieser Städte haben mehr als 250 000 Einwohner – 135 zwischen 50 000 und 250 000 Einwohner.

Weitere Erhebungsrunden fanden etwa im Drei-Jahres-Rhythmus zu den Bezugsjahren 2004, 2007 und 2010 statt. Die Zahl der teilnehmenden Städte und die der erhobenen Merkmale wuchs von Runde zu Runde an. Erhoben wurden jeweils demografische, wirtschaftliche, soziale, kulturelle und umweltspezifische Merkmale auf Gesamtstadtebene. Ergänzt wurden sie durch eine Auswahl aus dem Merkmalsprogramm für die Larger Urban Zones (LUZ) und die Subcity Districts (SCD) der UA-Städte.

Organisation:

Seit Urban Audit II liegt die Federführung europaweit bei EuroStat und in den EU-Mitgliedsstaaten bei den jeweiligen nationalen Statistischen Ämtern. In der Bundesrepublik wird die Datensammlung von der KOSIS-Gemeinschaft Urban Audit in Zusammenarbeit mit dem Statistischen Bundesamt und den Statistischen Landesämtern organisiert und durchgeführt. Hierzu werden bereits vorhandene amtliche oder halbamtliche Datenquellen genutzt. Daten aus Sonderaufbereitungen und Schätzungen (u. a. aus dem Mikrozensus) fließen ebenfalls mit ein, so dass ein Großteil der Datenanforderungen der EU erfüllt werden kann. Abstimmungen finden in der Lenkungsgruppe und der Mitgliederversammlung der KOSIS-Gemeinschaft sowie über das Netzwerk Stadt- und Regionalstatistik statt. Die betreuende Stelle war bis 2013 im Amt für Stadtforschung und Statistik in Nürnberg angesiedelt – sie wurde jüngst an die Kommunale Statistikstelle der Stadt Mannheim übergeben.

aktuell:

Aufgrund der wachsenden Bedeutung städtebezogener EU-Politik wurde Urban Audit inzwischen zu einer ständigen Aufgabe im Europäischen Statistischen System (ESS) erhoben. Der Katalog enthält Geofach- und Geobasisdaten. Mit der Erhebungsrunde zum Urban Audit V (Projektlaufzeit 2011 bis 2013) wurde das Erhebungsprogramm gestrafft und der zeitliche Turnus umgestellt. Europaweit soll nun ein Kerndaten-

bestand (57 Variable) jährlich erhoben werden, während der Gesamtdatenkatalog für Gesamtstadtebene (338 Variable) und LUZ (184 Variable) alle drei Jahre abgefragt wird. Die Daten der SCD (46 Variable) sollen gar nur noch alle zehn Jahre – gekoppelt an die Zensusjahre – für Großstädte mit mehr als 250 000 Einwohnern geliefert werden. Zugleich wird allen bisherigen UA-Städten die Möglichkeit gegeben, ihre SCD neu abzugrenzen. Um eine gewisse Informationstiefe zu gewährleisten, sollten pro Stadt mindestens zehn SCD ausgewiesen werden. Im Durchschnitt sollen somit etwa 25 000 Menschen zu einem SCD zählen.

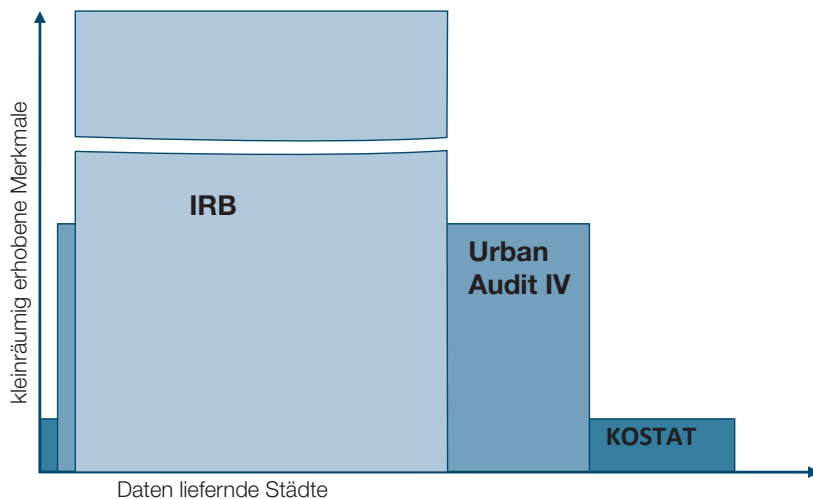
Durch Abstimmung der Städteauswahl mit den Gebietskategorien der amtlichen Stichprobenerhebungen – von der Arbeitskräftestichprobe bis zu SILK – wollen DG Regio und EuroStat einerseits das Spektrum städtebezogener Informationen erweitern und andererseits aktuelle Aussagen aus den Stichproben mit gebietlich differenzierten Aussagen aus dem Urban Audit kombinieren.

2.4 Zwischenfazit

Wenn wir die drei Kataloge hinsichtlich Zahl der Daten liefernden Städte und Zahl der Merkmalsausprägungen für kleinräumige Raumeinheiten unterhalb der Gesamtstadtebene zueinander in Beziehung setzen, stellt sich ihr Verhältnis bislang etwa wie in Abbildung 2 dar.

(7) Problemen der Übertragbarkeit dieser Gütekriterien auf andere Kontexte als den einer Messung werden von Ines Steinke (1999: 131–294) diskutiert.

Abbildung 2
Kataloge mit kleinräumigen Daten deutscher Kommunen nach Anzahl der Einzelmerkmale und Anzahl der Daten liefernden Städte



Quelle: eigene Darstellung

3 Kriterien für die Beurteilung von Datenqualität

Um die Qualität der Datensammlungen zu diskutieren, steht eine große Zahl von Kriterien zur Verfügung, deren Bedeutung vom Kontext und der Verwendung der Daten bzw. Informationen abhängt. Am bekanntesten sind innerhalb der Methoden empirischer Sozialforschung die traditionellen Gütekriterien für eine Messung bzw. einen Test: Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit).⁷ Recherchiert man heutzutage das Schlagwort „Datengüte“, wird meistens Datum mit Information gleichgesetzt und die Diskussion ist geprägt von Artikeln, die sich mit der Qualität von – häufig kommerziell genutzten – Datenbankinhalten befassen. So lassen sich für Datenqualität laut Richard Wang und Diane Strong (1996) auf Grundlage einer Befragung von ITNutzern die in Tabelle 1 aufgeführten Kriterien unterscheiden:

Problematisch wird von uns u. a. beurteilt, dass in diesem die Denkweise der Informatik spiegelnden System durch den durchgängigen Gebrauch des Begriffs „Information“ verschiedene Phasen empirischer Prozesse vermischt werden, die unterschiedlichen Anforderungen folgen sollten. So findet der von der Deutschen Gesellschaft für Informations- und Datenqualität (DGIQ) benutzte Informationsbegriff in der Raum- und Stadtbeobachtung des BBSR sein Pendant in den zu unterscheidenden Begriffen Daten, Indikatoren und Informationen. Dahinter stehen unterschiedliche Geltungsbereiche der Begriffe:

- Daten (vom lateinischen dare = geben, datum = das Gegebene) sind logisch gruppierte Einheiten, die zwischen Systemen übertragen werden oder auf Systemen gespeichert sind. Semiotisch betrachtet sind Daten potenzielle Informationen. Auf soziale Tatsachen und Prozesse bezogene Daten werden mittels Methoden empirischer Sozialforschung erhoben, analysiert und interpretiert – z. B. die Datenerhebung erfolgt mittels der Instrumente Sammlung, Beobachtung oder Befragung.

Tabelle 1

Kriterien für die Prüfung von Datenqualität nach Wang und Strong (1996) in Übersetzung der Deutschen Gesellschaft für Informations- und Datenqualität e.V. – DGIQ (2007)

Kriterien von Wang & Strong, 1996	Kriterien der DGIQ, 2007	Beschreibung laut DGIQ
Accessibility	Zugänglichkeit	Informationen sind zugänglich, wenn sie anhand einfacher Verfahren und auf direktem Weg für den Anwender abrufbar sind.
Appropriate Amount of Data	Angemessener Umfang	Informationen sind von angemessenem Umfang, wenn die Menge der damit verfügbaren Information den gestellten Anforderungen genügt.
Believability	Glaubwürdigkeit (Aufmachung, Marketing)	Informationen sind glaubwürdig, wenn Zertifikate einen hohen Qualitätsstandard ausweisen oder die Informationsgewinnung und -verbreitung mit hohem Aufwand betrieben werden.
Completeness	Vollständigkeit (incl. Pünktlichkeit)	Informationen sind vollständig, wenn sie nicht fehlen und zu den festgelegten Zeitpunkten in den jeweiligen Prozess-Schritten zur Verfügung stehen.
Concise Representation	Übersichtlichkeit	Informationen sind übersichtlich, wenn genau die benötigten Informationen in einem passenden und leicht fassbaren Format dargestellt sind.
Consistent Representation	Einheitliche Darstellung (Format, Layout, Wertevorrat)	Informationen sind einheitlich dargestellt, wenn die Informationen fortlaufend auf dieselbe Art und Weise abgebildet werden.
Ease of Manipulation	Bearbeitbarkeit	Informationen sind leicht bearbeitbar, wenn sie leicht zu ändern und für unterschiedliche Zwecke zu verwenden sind.
Free of Error	Fehlerfreiheit (Widerspruchsfreiheit, Genauigkeit)	Informationen sind fehlerfrei, wenn sie mit der Realität übereinstimmen.
Interpretability	Eindeutige Auslegbarkeit (geeignete Sprache, Symbole)	Informationen sind eindeutig auslegbar, wenn sie in gleicher, fachlich korrekter Art und Weise begriffen werden.
Objectivity	Objektivität (abhängig vom Daten generierenden Prozess)	Informationen sind objektiv, wenn sie sachlich und wertfrei sind.
Relevancy	Relevanz (Verbesserung der Zielerreichung)	Informationen sind relevant, wenn sie für den Anwender notwendige Informationen liefern.
Reputation	Hohes Ansehen	Informationen sind hoch angesehen, wenn die Informationsquelle, das Transportmedium und das verarbeitenden System im Ruf einer hohen Vertrauenswürdigkeit und Kompetenz stehen.
Timeliness	Aktualität	Informationen sind aktuell, wenn sie die tatsächliche Eigenschaft des beschriebenen Objektes zeitnah abbilden.
Understandability	Verständlichkeit	Informationen sind verständlich, wenn sie unmittelbar von den Anwendern verstanden und für deren Zwecke eingesetzt werden können.
Valueadded	Wertschöpfung	Informationen sind wertschöpfend, wenn ihre Nutzung zu einer quantifizierbaren Steigerung einer monetären Zielfunktion führen kann.

• Indikatoren (vom lateinischen *indicare* = anzeigen) sind empirisch feststellbare Sachverhalte, die stellvertretend für ein theoretisches Konzept oder einen theoretischen Begriff betrachtet werden. Um in diesem Sinne als Hinweis auf etwas anderes verwendet werden zu können, ist eine Operationalisierung notwendig und deren Regel ist zu veröffentlichen. Damit bilden Indikatoren Analyseinstrumente, mit denen in den Gesellschaftswissenschaften Lebensqualität oder Gesamtzustand und Entwicklungsvorgänge einer Gesellschaft quantitativ ermittelt und mit anderen Gesellschaften verglichen wer-

den. Anwendungen liegen in der Sozialberichterstattung oder der Begründung und Verbesserung wohlfahrtsorientierter politischer Planung.⁸

• Informationen (vom lateinischen *informare* = Gestalt geben) werden im allgemeinen Sprachgebrauch mit Kenntnissen oder Wissensinhalten, mit Bedeutung (semantische Ebene) gleichgesetzt. In Form situativ ausgewählter Informationen werden Ergebnisse von als relevant angesehenen Erfahrungsprozessen weitergegeben. In den verschiedenen wissenschaftlichen Disziplinen wird der Begriff sehr unterschiedlich

(8)
Zum Indikatorensystem der Laufenden Raum- und Stadtbeobachtung des BBSR siehe Heft 8/9.1978 der Informationen zur Raumentwicklung, dort insbesondere den Artikel von Hans-Peter Gatzweiler: 699ff.

verwendet. Als (syntaktisches) Grundprinzip ist allen disziplinären Perspektiven zuzueigen, dass Informationen auf Unterscheidbarkeit beruhen. Als Handlungsgrundlage brauchen sie zugleich Anteile von Wiedererkennbarkeit und Neuigkeitsgehalt. Informationen sollen Wissen vermehren, Ungewissheit verringern, übertragbar sein und im Prozess des Informierens den Empfänger bzw. das aufnehmende System verändern. Im Zusammenhang mit wissenschaftlicher Politikberatung kommt es insbesondere auf die Passgenauigkeit (Teil der pragmatischen Ebene) einer Information für den Empfänger an, damit sie Wirkung entfalten kann. Somit sind aus dem raumbezogenen Informationssystem des BBSR heraus entwickelte Informationen immer auch Interpretationen und mit Einschätzungen bzw. Bewertungen versehen. Erst diese machen

die weitergeleiteten oder veröffentlichten Informationen zu raumbezogenen bzw. zu politisch relevanten Aussagen (funktionaler Aspekt).

In diesem Sinne beziehen sich die oben aufgeführten Qualitätskriterien der DGIQ

- Accessibility, Completeness, Concise representation, Consistent representation, Free of error, Objectivity sowie Timeliness auf Daten bzw. eine Datensammlung,
- Appropriate amount of data, Ease of manipulation, Interpretability, Timeliness sowie Relevancy auf Indikatoren

und

- Accessibility, Believability, Relevancy, Reputation, Timeliness, Understandability sowie Valueadded auf Informationen.

Tabelle 2
Gütekriterien nach ESS Standards for Quality Reports (Eurostat 2009)

Eurostat-Standards	Identifications	Kurzbeschreibungen (freie Übersetzung: Gutfleisch und Sturm)
Relevance	Rate of available statistics	Relevanz bezeichnet das Ausmaß, in dem statistische Ergebnisse und deren Analyse die Bedarfe der Nutzer abdecken können.
Accuracy	<ul style="list-style-type: none"> • Coefficient of variation • Rate of overcoverage • Edit failure rate • Unit response rate • Item response rate • Imputation rate • Number of mistakes made, by type • Average size of revisions 	Genauigkeit von Daten bzw. statistischer Ergebnisse und deren Analyse bezeichnet die Wahrscheinlichkeit, mit der die Schätzwerte bzw. die abgeleiteten Aussagen die wahren Werte bzw. Tatsachen abbilden.
Timeliness and Punctuality	<ul style="list-style-type: none"> • Time lag between end of reference period and date of first / provisional results – final results • Punctuality of publication 	Aktualität von Daten bzw. statistischer Analysen bezieht sich auf die Zeitspanne zwischen dem bezeichneten Ereignis und der Verfügbarkeit der Daten bzw. Analysen – Pünktlichkeit bezieht sich auf die Verzögerung zwischen Freigabezeitpunkt der Daten oder Analysen und Veröffentlichungs- bzw. Übergabetermin an die Nutzer.
Accessibility and Clarity	<ul style="list-style-type: none"> • Number of subscriptions / purchases of each of the key paper reports • Number of accesses to online databases • Rate of completeness of metadata 	Zugänglichkeit und Klarheit verweisen auf die Handhabbarkeit der Daten, Statistiken oder Analysen für Nutzer aufgrund geeigneter Metadaten, Handbücher oder Begleitinformationen.
Coherence and Comparability	<ul style="list-style-type: none"> • Lengths of comparable time series • Asymmetries of statistics mirror flows 	Kohärenz zweier Datensätze oder Analysen bezieht sich auf das Ausmaß, in dem die Datenerhebungen bzw. die Auswertungen vergleichbaren Konzepten und Methoden folgen – Vergleichbarkeit bezeichnet einen Spezialfall von Kohärenz, in dem dieselben Merkmale / Variablen für Vergleiche zwischen Regionen oder Zeitpunkten oder ... genutzt werden sollen.
Assessment of User Needs and Perceptions	<ul style="list-style-type: none"> • User satisfaction index • Length of time since most recent user satisfaction survey 	Der formulierte wie noch nicht formulierte Bedarf, Fragen und Erwartungen von internen wie externen Nutzern bilden die Leitlinien der (ESS-)Arbeit.
Performance, Cost and Respondent Burden	<ul style="list-style-type: none"> • Annual operational cost, with break-down by major cost components • Annual respondent burden in hours and/or financial terms 	Ressourcen müssen effektiv genutzt werden. Die Belastungen der Datenlieferanten und Analysten sollten dem Bedarf der Nutzer angemessen sein. Im Laufe der Zeit sollte der Aufwand sinken.
Confidentiality, Transparency and Security		Die Anonymität der Merkmalsträger, ein hinreichender Datenschutz und die Verwendung der Daten ausschließlich für wissenschaftliche Zwecke müssen garantiert sein. Wissenschaftliche Unabhängigkeit ist zu respektieren und Nutzer sind gleich zu behandeln.

In Bezug auf die hier betrachteten Kataloge kleinräumiger Daten der Kommunalstatistik sind hauptsächlich die ersten beiden Kriteriengruppen von Interesse – es geht um die Qualität der Daten an sich und um die Qualität der Indikatoren, die das dahinterstehende theoretische Konzept repräsentieren. Sie sind die Grundlage für die Aufnahme bestimmter (Aggregat-)Merkmale in die Kataloge.

Um die Qualitätskriterien für eine Beurteilung der vorliegenden Kataloge mit Kriterien aus einer anderen – eher transdisziplinären Praxis stammenden – Quelle abzugleichen bzw. zu modifizieren, ziehen wir eine weitere Kriterienliste von Eurostat zu Rate. Deren Richtlinien für qualitätvolle Statistik (2009) sehen die in Tabelle 2 aufgeführten Standards bzw. Gütekriterien vor.

Für die weitere Betrachtung der drei Kataloge beschränken wir uns auf die kleinräumigen kommunalstatistischen Daten – also auf Geofachdaten und nicht auf Geobasisdaten, die im KOSTAT-Katalog und unter Einschränkung im Urban Audit, jedoch nicht in der IRB abgelegt sind. Aus den beiden vorgestellten Kriterienlisten für Datenqualität formulieren wir folgende Beurteilungsaspekte, die sich stark an die Eurostat-Standards anlehnen und die DGIQ-Kriterien weitgehend subsumieren.

Im Zusammenhang mit den methodologischen Vorüberlegungen bzw. als konzeptionelle Rahmung jeder wissenschaftlichen Arbeit ist nachzudenken über

- Relevanz

Bei der Datenerhebung bzw. der Zusammenstellung der Kataloge ist zu achten auf

- Genauigkeit
- Aktualität
- Ressourcen

Bei der Datenverwendung wird von Nutzerinnen und Nutzern besonderer Wert gelegt auf

- Eindeutigkeit
- Korrektheit

Und schließlich ist für zahlreiche Fragestellungen eine Verknüpfung verschiedener Datensätze wegen der unterschiedlichen Inhalte sinnvoll. Dies ist nur möglich bei ausreichender

- Kohärenz

Die ersten sechs Kriterien werden im Folgekapitel 4 reflektiert, während dem letzten Aspekt im Kapitel 5 etwas mehr Aufmerksamkeit gewidmet wird.

4 Beurteilung der Kataloge

4.1 Relevanz

Die Relevanz eines Datensatzes wird sowohl vom Entdeckungs- als auch vom Verwendungszusammenhang her gedacht. Es geht sowohl um die Fragen und Erwartungen als auch um den Bedarf der aktuellen wie potenziellen Nutzerinnen und Nutzer. Methodologisch gehören dazu Überlegungen zur Gegenstandsangemessenheit bzw. Indikation einerseits wie zur Limitation bzw. Restriktion in diesem Falle amtlicher Daten andererseits.

Im Falle der Kataloge mit kleinräumigen kommunalstatistischen Daten richtet sich der Blick relativ schnell von den idealtypischen wissenschaftlichen Fragestellungen auf die praktische Begrenztheit der möglichen Operationalisierungen aufgrund der stadt-spezifisch sehr unterschiedlich entwickelten Statistiken bzw. Dateninventare.

- Auf Grundlage der kommunalen Melderegister enthalten alle drei Kataloge demografische Variablen, die die Verteilungen von Geschlecht, von Altersgruppen und von Nationalitäten in unterschiedlicher Tiefe abbilden.
- IRB und UA enthalten darüber hinaus auch – über die Städte weitergeleitete – Daten der Bundesanstalt für Arbeit (BA) und der Sozialämter zu sozialversicherungspflichtig Beschäftigten, Arbeitslosen und Leistungsempfängern. Weiter werten die Städte für die Datenlieferungen an diese Kataloge ihre Register hinsichtlich der Wahlbeteiligung, der Struktur der Privathaushalte, dem Wohnungsbestand und der Flächennutzung aus.
- Die IRB sammelt zudem Familienstandsdaten, Bewegungsdaten – sowohl natürliche Bevölkerungsbewegungen als auch Umzüge und Wanderungen nach Alter – und Daten zur Wohndauer.
- UA sammelt auch auf SCD-Ebene zusätzlich Daten zum durchschnittlichen Einkommen der Privathaushalte, zu Bildung, zu Sterbefällen und zu Straftaten.

Alle drei Kataloge haben ihre Wurzeln in den Anforderungen, die wissenschaftliche Politikberatung stellt. Der KOSTAT-Katalog zielt zudem auch auf Nutzerinnen und Nutzer aus sozialwissenschaftlicher Grundlagenforschung wie auf privatwirtschaftlich agierende Geodatenhalter. Zur Beurteilung der Relevanz unter Fokussierung des Verwendungszusammenhangs wird in keinem Fall eine Befragung der Nutzerinnen und Nutzer durchgeführt. So kann dieser Aspekt der Relevanz nur mehr oder weniger indirekt geprüft werden:

- Bei KOSTAT lassen die Ankäufe des Datensatzes durch vor allem privatwirtschaftlich arbeitende Dritte auf Bedarf schließen.
- Bei der IRB spiegelt sich der Nutzen unter anderem in den Anfragen nach Vorträgen, wissenschaftlichen Beiträgen in Fachpublikationen, inhaltlichen Analysen für das Bundesministerium für Verkehr, Bau und Stadtentwicklung (BMVBS) oder für andere (standes-)politische Gremien und Institutionen. Weiterhin nutzen an der IRB beteiligte Städte Teile der Analysen für kommunalpolitische Entscheidungen. Auf dieser Grundlage wird der Merkmalskatalog regelmäßig in den Mitgliederversammlungen des Kooperationsnetzwerks diskutiert und gegebenenfalls erweitert oder gekürzt.
- Bei UA kann anhand der Pretests, die mit den bisherigen Daten seitens EuroStat oder bei nationalen Analysen im Hinblick auf die Lissabon-Strategie (Stadtentwicklungsdimensionen: Wettbewerbsfähigkeit, Soziale Kohäsion) vorgenommen werden, auf die Eignung geschlossen werden. Die daraus gewonnenen Erfahrungen wurden seitens der DG Regio zur aktuellen Umstrukturierung genutzt.

Bislang können die Kataloge mit kleinräumigen innerstädtischen Daten den Bedarf der aktuellen wie potenziellen Nutzerinnen und Nutzer aus Politik und Wissenschaft nur begrenzt abdecken. Ein Teil der möglichen Kataloginhalte beruht auf kommunalen Registern, die in möglichst vielen der jeweils am Katalog beteiligten Städte vergleichbar vorhanden und erschlossen sein müssen. Deshalb wird der so gewinnbare Datendfundus beschränkt bleiben. Andere Datenquellen, die eine kleinräumige Zuordnung erlauben könnten, werden von den verant-

wortlichen Bundesländern oder anderen Behörden derzeit für die Städte (noch) nicht freigegeben. So stehen beispielsweise für kleinräumig stadtvergleichende Analysen weder eine Steuerstatistik noch Umweltdaten (beispielhaft für Berlin: Klimeczek 2012, 2013) zur Verfügung. Zugleich kann die Relevanz der bereits vorhandenen Inhalte und Analysemöglichkeiten im Bewusstsein der politischen wie der wissenschaftlichen Öffentlichkeit noch deutlich verbessert werden. Dazu fehlt es den Katalog führenden Stellen unter anderem jedoch an personellen – teils auch datentechnischen – Kapazitäten (siehe Kapitel 4.4: Ressourcen). Trotz dieser Rahmenbedingungen können einige Städte im Rahmen von Open Data ihre kleinräumigen Daten nun im Internet mit Hilfe von Strukturdatenatlanten kostenlos bereitstellen. Hier seien als Beispiel Bremen, Frankfurt, Hamburg, München, Nürnberg und Stuttgart erwähnt.⁹

4.2 Genauigkeit (auch: Richtigkeit, Zuverlässigkeit)

Als auf den Prozess und das Ergebnis der Datenerhebung (Sammlung kleinräumiger innerstädtischer Statistiken) bezogenes Gütekriterium soll Genauigkeit hier alle Aussagen versammeln, die die Güte der Abbildung von gesellschaftlichen Tatsachen (Realität) in den Datenkatalog beurteilen. Dazu zählen hier Eindeutigkeit der Definition der Aggregatmerkmale, Einfachheit und Fehlerfreiheit bei Dateneingabe und Datenübertragung, Vollständigkeit, Schätzung oder Festlegung einzelner Werte bzw. Interpolationen innerhalb von Zeitreihen, Anpassung an Gebietsstandsänderungen im Zeitverlauf und Anpassung an Änderungen der administrativen Grenznetzwerke im Zeitverlauf.

Der Prozess des Einsammelns der kleinräumigen Kommunaldaten startet jeweils in der Daten sammelnden Stelle. Dort werden die Excel-Tabellen für die einzelnen Aggregatmerkmale erstellt mit den genauen Beschreibungen und Ausprägungen der Merkmale im Tabellenkopf. Alternativ erprobt Urban Audit seit kurzem die Erhebung mit einer Internet-Eingabemaske über das DUVA-System. In den Daten liefernden Kommunen durchläuft die Datenbereitstellung häufig mehrere Büros, wobei von den Daten sammelnden Stellen nicht beurteil-

(9) Aufgrund des geringen technischen Aufwandes und der einfachen Downloadmöglichkeiten der Daten nutzen alle hier aufgeführten Städte den sogenannten *InstantAtlas*.

bar ist, welche Qualitätskontrollen bereits an den Quellorten erfolgen oder ob alle relevanten Metadaten angeführt werden. Eine dokumentierte Kontrolle erfolgt dann erst beim Zusammenspielen der kommunalen Datensätze zu einem Jahrgangsdatsatz und gegebenenfalls bei dessen Verknüpfung mit den bereits existierenden früheren Jahresscheiben der Zeitreihe.

Da die hier betrachteten Kataloge für leicht unterschiedliche Zwecke sammeln, wird bei der Zusammenstellung entsprechend unterschiedlich Wert auf Genauigkeit gelegt:

- Für den KOSTAT-Katalog liegt der Schwerpunkt darin, von möglichst vielen Städten einen festen Datenstamm möglichst zeitnah für Interessierte zur Verfügung zu stellen. Dadurch sind für die Daten eines Jahres nur oberflächliche Plausibilitätsprüfungen möglich, bei denen allenfalls grobe Fehler – wie fehlende oder außerhalb des Wertebereiches liegende Daten – auffallen. Auch können seitens der Daten sammelnden Stelle die Jahresdatensätze nicht zeitreihenfähig zusammengefügt werden. Positiv ist zu vermerken, dass sich im Verlauf der Jahre seitens der Daten liefernden Kommunen die Qualität der Datenlieferung allein durch die wiederholte Praxis verbessert hat (siehe dazu auch den Artikel von Jürgen Göttsche-Stellmann in diesem Heft).

- Für die IRB durchläuft ein neu in den Katalog aufzunehmender Jahresdatensatz verschiedene standardisierte Plausibilitätsprüfungen (siehe dazu den Artikel von Antje Güleş in diesem Band), um zumindest augenscheinliche Fehler oder Widersprüche korrigieren zu können. Allerdings sind so nur selten falsch eingegebenen Daten zu finden, die innerhalb gültiger Wertebereiche liegen oder logisch konsistent erscheinen. So fallen regelmäßig Fehler erst bei inhaltlichen Analysen seitens des BBSR oder anderer Nutzerinnen und Nutzer auf, die – sofern sie rückgemeldet werden – im Nachhinein zu einer Korrektur auch älterer Daten führen. Probleme bereiten zudem Gebietsstandsänderungen in den Kommunen: Sofern dabei keine Änderung des innerstädtischen Lagetyps erfolgt, blähen „neue“ Stadtteile zwar den Datensatz auf, tangieren jedoch nicht die Zeitreihenanalysen auf Lagetypenebene.

- Für Urban Audit werden in einem ersten Schritt die in Deutschland erhobenen Daten in der betreuenden Stelle einer standardisierten Plausibilisierung unterzogen. Auch bei diesem nationalen Ausschnitt des Katalogs ist festzustellen, dass die kommunale Praxis des Sammelns die Datenqualität über die Jahre verbessert. Weiter müssen die kommunalen Daten in einem zweiten Schritt jedoch an die Bevölkerungsfortschreibung des Bundes und der Länder angepasst werden, um keinen Widerspruch zu den anderen EU-Statistiken der Bundesrepublik zu erzeugen. Diese an die EU gelieferten Daten werden seitens EuroStat weiteren Plausibilisierungen unterzogen. Damit wird der deutsche Urban-Audit-Datensatz aufgrund unterschiedlicher Bezugssysteme quasi gedoppelt. Wie richtig die an EuroStat weitergeleiteten Daten sind, kann von den liefernden Kommunen wegen der Umschätzung kaum mehr beurteilt werden. Dadurch sind viele der UA-Daten für innerdeutsche Stadtvergleiche eher irrelevant.

Bei der Arbeit mit den kleinräumigen innerstädtischen Aggregatdaten sind also Einflussfaktoren und Referenzrahmen zu berücksichtigen – im Falle der UA-SCD gibt es sogar zwei parallele Datensätze, die in Bezug zum Referenzsystem beide richtig sind. Genauigkeit und Zuverlässigkeit kann sich folglich nur auf die Konsistenz der Dateneingabe und Datenübertragung beziehen – und gegebenenfalls auch auf die Umschätzung. Hinsichtlich unterschiedlicher Merkmalsdefinitionen oder Gebietsstandsänderungen ist auf entsprechend hinterlegte Meta- und Paradata¹⁰ zu achten.

4.3 Aktualität (auch: Pünktlichkeit)

Zu diesem Qualitätskriterium zählen der eindeutige Referenzzeitpunkt der Daten und eine eindeutige Periodizität sowie eine möglichst kurze Zeitspanne zwischen dem Referenzzeitpunkt und der Datenbereitstellung bzw. der Datenanalyse. Letzteres ist gegen die Gesamtgenauigkeit abzuwägen.

- IRB und KOSTAT erheben ihre Daten jährlich jeweils zum Referenzdatum 31.12. des vergangenen Jahres (BA-Daten zum 15.09. des Jahres) bzw. bei den Bewegungsdaten für die Zeitspanne des vergangenen Kalenderjahres. Bei UA wurden die SCD-Daten bislang im Dreijahresrhythmus erhoben – zukünftig im Zehnjahresrhythmus.

(10) Die Sozialwissenschaften unterscheiden Paradata und Metadaten.

Metadaten sind im Prinzip „Daten über Daten“. Im Falle von Geodaten stellen sie „dem Nutzer eine strukturierte Beschreibung der eigentlichen Geodaten und Geodienste zur Verfügung und sind somit ein elementarer Baustein einer funktionierenden Geodateninfrastruktur. ... Metadaten sollten der breiten Öffentlichkeit zur Verfügung stehen. Die Bereitstellung der Metadaten erfolgt über eine standardisierte Schnittstelle, den sogenannten Katalogdienst (CSW – Catalogue Service Web)“ (www.geodaten.niedersachsen.de). Im Falle der hier behandelten Aggregatdaten zählen neben dem Zeitpunkt der Einführung einer Zweitwohnsitzsteuer oder Gebietsstandsänderung insbesondere die Schätzalgorithmen, mit denen fehlende Daten in einer Zeitreihe ersetzt werden oder mit denen Mikrozensusdaten auf Kreisebene umgerechnet werden, zu den Metadaten.

Paradata entstehen als Nebenprodukte des Datenerhebungsprozesses. Beispiele aus diesem Zusammenhang sind die Zeit, die zur Beantwortung einer Frage bei einer CAPI- oder CATI-Befragung benötigt wird (durch Software erhoben), oder Abbruch- bzw. Verweigerungsgründe (durch Interviewer erhoben). Auch Kontextdaten – wie zum Beispiel Raumtypen – werden in manchen Erhebungssituationen als Paradata verstanden. Bei den hier behandelten Aggregatdatensätzen könnten zum Beispiel die Datenlücken als Paradata verstanden werden.

- Die Bereitstellung der je neuesten Jahresscheibe erfolgt bei KOSTAT etwa im Oktober, bei der IRB etwa im Dezember jeden Jahres. Die IRB-Daten sind dann auch auf ihre Anschlussfähigkeit an die bestehenden Zeitreihen geprüft, während dies für Analysen mit den KOSTAT-Daten von den jeweiligen Nutzerinnen und Nutzern eigenständig vorgenommen werden muss. Der Bereitstellungsprozess der UA-Daten dauerte bislang deutlich länger, was durch die Umstellung auf einen kleinen Kranz jährlich zu liefernder Gesamtstadtdaten geändert werden soll. Bei den kleinräumigen SCD-Daten wird die Verfügbarkeit jedoch auch zukünftig länger auf sich warten lassen, da die Veröffentlichung der Zensusdaten, aus denen die Daten für die SCD gewonnen werden sollen, einer umfangreichen Datenaufbereitung bedarf.

Je nach gesellschaftlichem Themenfeld ändern sich die Gegebenheiten mehr oder weniger schnell. Wenn wissenschaftliche Analysen für Politikberatung einsetzbar sein sollen, müssen sie der Problemstellung entsprechend zeitnah zur Verfügung stehen. Gleichwohl sind zeitliche Verzögerungen nicht zu umgehen, wenn die Daten dem Kriterium der Genauigkeit in akzeptablem Maß genügen sollen.

4.4 Ressourcen

In einem Datenkatalog stecken konzeptionelle Kreativität, wissenschaftliches Know-How, (Arbeits-)Zeit, technische Infrastruktur und Geld vieler Personen wie öffentlicher oder privater Träger. Insofern hängen Umfang und Güte eines Katalogs stark von den vorhandenen Ressourcen bzw. Restriktionen ab. Über die Ressourcen nachzudenken lenkt den Blick auch auf das Zusammenwirken der Beteiligten in den verschiedenen Schritten der Datenerhebung, denn: Je geringer die Genauigkeit ist desto höher fällt der Bereinigungsaufwand aus.

Bei allen drei Katalogen durchläuft der Datensammel- und Datenaufbereitungsprozess mehrere Stationen, die bereits im vorhergehenden Abschnitt angesprochen wurden – und in jedem dieser Abschnitte kann es zu Ressourcenengpässen kommen:

- Personalmangel wird unter anderem dann offensichtlich, wenn eine Daten liefernde Kommune ihre Lieferung nicht im verabredeten Zeitrahmen realisieren kann, wenn sich bei den Daten sammelnden Stellen die Aufbereitung und das Zusammenspielen der eingehenden Datensätze verzögert oder wenn inhaltliche Analysen nur auf Sparflamme erfolgen.

- Mängel in der Datenhaltungsinfrastruktur treten auf, wenn die Datengenerierung und Datenverarbeitung nicht überall einem vergleichsweise neuen Standard folgen. Dann muss auf den diesbezüglich kleinsten gemeinsamen Nenner zurückgegriffen werden – was auf jeden Fall mit Zeitverlusten einhergeht und teils auch mangelhafte Internet-Präsenz zur Folge hat. Bei den betrachteten Katalogen heißt dies, dass die Daten mehrheitlich nach wie vor im CSV- bzw. Excel-Format erhoben und in einem Extraarbeitsschritt in ein anderes Datenhaltungsformat überführt werden müssen.

Die Gesamtheit der Kosten ist abzuwägen mit dem Nutzen bzw. Mehrwert, den Öffentlichkeit, Politik und Wissenschaft daraus jetzt und zukünftig ziehen können. Auch zukünftig anfallende Kosten für Analysen auf Basis eines existierenden Katalogs (siehe auch Kapitel 4.5: Eindeutigkeit) sind zu bedenken. Derzeit gibt es allerdings kein standardisiertes Verfahren, mit denen vorhandene oder neue Statistiken gegenüber den Kosten dieser Statistiken abgewogen und gerechtfertigt werden können. Dies gilt entsprechend für die Einschätzung, ob sämtliche Statistiken weiterhin benötigt oder eingestellt bzw. eingeschränkt werden sollten (siehe dazu Statistisches Bundesamt 2012: 3).

4.5 Eindeutigkeit

Da Datenkataloge i.d.R. nicht nur von der primär sammelnden Stelle verwendet werden, sollten sie in einer klaren Gliederung, gut dokumentiert und für alle potenziellen Nutzerinnen und Nutzer eindeutig interpretierbar vorliegen. Unter dem Teilaspekt der Zugänglichkeit ist zu klären, wo der Katalog geführt wird und welche Bedingungen bzw. Nutzungsregeln bei einem möglichen Zugriff auf die Daten gelten. Die Daten sollten in einem einheitlichen Format abgefasst

sein, das von unterschiedlichen Analyseprogrammen gleichermaßen gut eingelesen werden kann. Weiterhin muss die Datenstruktur so übersichtlich und verständlich sein, dass im Normalfall keine Fehlzuweisungen erfolgen können. Und schließlich ist der Katalog so ausführlich und verständlich zu dokumentieren, dass sich Dritte relativ schnell in den Datensatz und seine Verarbeitungsmöglichkeiten einfinden können. Zu den zu dokumentierenden Metadaten gehört u. a. auch der Geltungsbereich der Merkmalsdefinitionen (siehe auch Kapitel 4.2: Genauigkeit). Methodologisch ist diese Kriterienebene hinsichtlich einiger Aspekte zu vergleichen mit dem Gütekriterium der intersubjektiven Nachvollziehbarkeit, denen qualitative Analyseschritte und damit erzeugte Forschungsergebnisse genügen müssen.

Für die drei Kataloge mit kleinräumigen Kommunaldaten ist dieses Gütekriterium gut erfüllt. Die Zugänglichkeit ist jeweils klar geregelt und Metadaten sind dokumentiert (soweit sie der Daten sammelnden Stelle zur Verfügung gestellt wurden). Alle Katalogauszüge stehen gegebenenfalls in mehreren Datenformaten (CSV, Excel, SPSS) zur Verfügung.

4.6 Korrektheit

Der Standard der Korrektheit soll hier sicherstellen, dass ein Datenkatalog rechtlich und ethisch korrekt zusammengestellt wurde und wird und so auch von allen – auch zukünftigen – Nutzerinnen und Nutzern zu verwenden ist.¹¹ Dazu gehört, dass dem Wohlergehen der durch die Daten repräsentierten bzw. durch die Analyseergebnisse betroffenen Personen gebührende Aufmerksamkeit gewidmet wird. Letzterem dienen z. B. Datenschutzbestimmungen¹² und Anonymisierungsregeln. Im Zusammenhang mit Korrektheit stehen weiterhin die Vertrauenswürdigkeit der Datenquellen und die Glaubwürdigkeit wie wissenschaftliche Unabhängigkeit der Katalogbetreiber. In der Arbeit mit Datensammlungen verpflichten sich Katalogbetreiber wie Nutzerinnen und Nutzer, auf Quellen und Datengrundlagen den wissenschaftlichen Gepflogenheiten entsprechend hinzuweisen.

Für alle drei Kataloge werden die Daten nach bestem Wissen und Gewissen der Daten sammelnden Stellen zusammengestellt. Die kommunikative Kontrolle des angestrebten objektiven, transparenten und sicheren Vorgehens findet in der AG KOSTAT, der IRB-Mitgliederversammlung oder der UA-Lenkungsgruppe bzw. der UA-Mitgliederversammlung statt. Da die Daten für Analysezwecke verbreitet an Dritte weitergegeben werden, gilt für den KOSTAT- und IRB-Katalog für gering besetzte Zellen eine Anonymisierungsregel.¹³

Zusammenfassend ist festzustellen, dass die Relevanz der Kataloge mit kleinräumigen innerstädtischen Daten für raumbezogene Wissenschaften wie für raumrelevante Politik hoch ist. Dass damit auch als brisant eingestufte Befunde erarbeitet werden können, ist im Effekt nicht nur förderlich für die Unterstützung dieser Datensammlungen. Insofern ist eine mehr oder weniger deutliche Ressourcenknappheit zu erkennen, was sich auf die Genauigkeit und Aktualität der Kataloge auswirkt. Hinsichtlich Eindeutigkeit und Korrektheit bieten die Kataloge den Nutzerinnen und Nutzern wiederum eine zeitgemäß gute Qualität.

5 Probleme der Kohärenz

Im Prinzip fragen sich alle Beteiligten immer mal wieder, warum die drei vorgestellten Kataloge für kleinräumige kommunalstatistische Daten deutscher Städte nicht zu einem zusammengefügt werden. Dem steht vor allem entgegen, dass sie für unterschiedliche Adressaten entwickelt und ausgebaut wurden und deshalb spezifische Charakteristika aufweisen. Trotzdem ist bei kleinräumigen Statistiken zu denselben Kommunen wünschenswert, dass sie widerspruchsfrei zusammenpassen bzw. in einem Referenzsystem eindeutig aufeinander zu beziehen sind. Im engeren Sinne geht es bei den drei Katalogen um Vergleichbarkeit bezogen auf

- Stichtag der Datenerhebung und zeitlichen Turnus,
- Zuschnitt der kleinräumigen Einheiten als Merkmalsträger,
- einheitliche Definition der Merkmale.

(11)
Siehe DGS & BDS (1992) oder DFG (1998).

(12)
Siehe BMJ (1990) und BfDI (2007).

(13)
Wenn der Datensatz oder Auszüge aus demselben an Dritte weitergegeben wird/werden, wird zuvor eine ABS-perturbation durchgeführt. Angelehnt an das auf eine Output-Tabelle aufsetzende Verfahren des Australian Bureau of Statistics (ABS) werden alle Zellbesetzungen, die 1, 2 oder 3 betragen, per Zufall durch 0 oder 4 ersetzt. Die externen Nutzerinnen und Nutzer werden darauf hingewiesen, dass sich durch dies Verfahren inkonsistente Randsummen ergeben.

5.1 Stichtag der Datenerhebung und zeitlicher Turnus der Sammlung

Als ein wichtiges Kriterium für die Vergleichbarkeit von Daten gilt der gemeinsame Stichtag, der Zeitpunkt, auf den sich die statistische Erfassung der Merkmale bezieht. Der Stichtag wird bei den hier vorgestellten Datensammlungen durch die Leitvariable Bevölkerung bestimmt, deren Datenquelle das Einwohnermelderegister darstellt. Um vergleichbare statistische Auswertungen und den Einsatz standardisierter Auswertungsprogramme in den Städten zu ermöglichen, ist ein einheitlicher „Statistikdatensatz Bevölkerungsbestand“ vom Deutschen Städtetag und dem KOSIS-Verbund festgelegt worden, so dass von einem standardisierten Datensatz gesprochen werden kann. Standard ist hier der 31.12. eines Jahres. Bei allen drei Datensammlungen stehen zu diesem Stichtag Bevölkerungsdaten zur Verfügung. Die darüber hinaus gesammelten Merkmale stehen – sofern es sich nicht um Verlaufsdaten handelt – ebenfalls zu diesem Stichtag bereit. Eine wichtige Ausnahme sind die Merkmale der Bundesagentur für Arbeit. Deren Datenstand wird, um Saisonschwankungen auszugleichen, abweichend gewählt. Er kann jedoch zu jedem beliebigen Datum generiert werden. Damit stehen für alle Merkmale der drei Datensammlungen grundsätzlich die jeweils gleichen Stichtage zur Verfügung.

Die häufig geforderte Datenaktualität spielt in diesem Zusammenhang eine eher untergeordnete Rolle, da es sich bei allen drei Datensammlungen um Strukturdaten und nicht um Konjunkturdaten handelt.

Anders sieht die Lage bei der Sammlung der Daten aus. Diese erfolgt bei allen drei Sammlungen im jährlichen Turnus bei den Städten. Die Schwierigkeit für die Städte besteht darin, dass der Abgabetermin an die sammelnde Stelle zu unterschiedlichen Zeitpunkten erfolgt. Während für die AG KOSTAT bereits im April des Folgejahres Daten bereitgestellt werden müssen, endet die Abgabefrist für die IRB-Daten erst Ende Oktober des Folgejahres. Die Geschäftsstelle von Urban Audit sammelt die Daten eineinviertel Jahre später ein. So besteht für die/den einzelne/n Sachbearbeiter/in häufig die Schwierigkeit der Zuordnung der Merkmale zu den einzelnen Datensammlungen gerade auch unter der Einbeziehung

der steigenden Anzahl von umfangreichen Datenanfragen von privatwirtschaftlicher Seite. Auch unter dem Aspekt der Kostenminimierung und des weiteren Personalabbaus in den Statistikstellen sollte hier zukünftig eine Annäherung erfolgen.

5.2 Zuschnitt der kleinräumigen Einheiten als Merkmalsträger

Die administrative Gliederung der Bundesrepublik ist bis zur Gemeindeebene hierarchisch aufgebaut. Den hoheitlichen Aufgaben entsprechend sind Bund, Länder, Regierungsbezirke bzw. Regionen, Landkreise bzw. kreisfreie Städte und Gemeinden zu unterscheiden. Unterhalb der Gemeindeebene ist die Untergliederung in weitere Einheiten nicht einheitlich definiert. Sie ist vielmehr abhängig von der Gemeindeordnung sowie der Größe und funktionalen Bedeutung der Stadt. Bei allen Großstädten liegt eine kleinräumige Gliederung vor, ein statistisches Raumbezugssystem. Es teilt die Stadt in verschiedene Einheiten auf, die hierarchisch aufgebaut sind. Kleinste Einheit stellt immer die Adresse (Straße und Hausnummer) dar. In Frankfurt a. M. als Beispiel folgen darauf aufbauend der Block und dessen Blockseite, der Wahlbezirk, der Stadtbezirk, der Stadtteil und schließlich der Ortsbezirk.

Tabelle 3
Das statistische Raumbezugssystem der Stadt Frankfurt a. M.

Raumbezug	Anzahl Einheiten
Adressen	rd. 87 000
Blöcke	rd. 6 200
Wahlbezirke	rd. 460
Stadtbezirke	121 (3-Steller)
Stadtteile	47 (2-Steller)
Ortsbezirke	16 (1-Steller)

Dieser hierarchische Aufbau ist in allen Städten ähnlich, auch wenn die Begrifflichkeiten in den einzelnen Städten verschieden sein können. Für die drei Datensammlungen werden von den Städten – bleibt man bei der Frankfurter Bezeichnung – Merkmale für die räumliche Ebene der Stadtbezirke (bevorzugt für KOSTAT) bzw. die nächst höhere Ebene der Stadtteile (bevorzugt für IRB) geliefert. Zwischen diesen beiden Ebenen der sogenannten 2- und 3-Steller hat jede Stadt eine Umstiegs-Referenz (Metadatum), so dass prinzipiell aus KOSTAT in

IRB umgerechnet werden kann. Für Urban Audit wurde von den Städten selbst eine neue Ebene generiert, die sogenannten UA-Districts, die nicht alle deutschen UA-Städte auf Basis ihres zuvor existierenden statistischen Raumbezugssystems generiert haben. Da bietet die von der DG Regio beabsichtigte Neugliederung jetzt die Chance einer Angleichung.

Ein direkter Vergleich auf Ebene der 1-, 2- und 3-Steller ist grundsätzlich schwierig. Zwar ähnelt sich der Aufbau der kleinräumigen Gliederung zwischen den Städten, jedoch sind die einzelnen Raumeinheiten im intrakommunalen Vergleich sehr heterogen. Die Verfahren zur kleinräumigen Abgrenzung wurden nie standardisiert. Beim Großteil der Städte sind es historische Gründe, wie z. B. Stadterweiterungen oder Eingemeindungen, die zur Bildung der Abgrenzungen geführt haben. Dies führt dazu, dass sich die Einheiten schon innerhalb der Gemeinden hinsichtlich Größe und Einwohnerzahl deutlich unterscheiden. So erstreckt sich die Spannweite innerhalb der Raumeinheit der Stadtbezirke in Frankfurt a.M. im Jahre 2009 zwischen rund 730 und rund 18 100 Einwohnerinnen und Einwohnern. Lange Konstanz in den Datenreihen ließen Veränderungen bei der Grenzziehung in der Vergangenheit generell kaum zu.

Durch die unterschiedlichen Größenklassen kann es bereits bei Analysen innerhalb einer Stadt zu Abweichungen, Verzerrungen und Nivellierungen kommen. Kleinräumige Vergleiche über mehrere Städte hinweg steigern die Schwierigkeiten und bedürfen umfangreicher Erfahrungen bei der Interpretation. Durch Kategorisierungen bzw. Zusammenfassungen ist es möglich, diese Problematik zumindest teilweise zu umgehen (vgl. auch Sturm 2010). Unter anderem dazu werden für die IRB – wie oben ausgeführt – innerstädtische Lagetypen gebildet (City, Cityrand, Innenstadtrand, Stadtrand). Die durchgeführten Vergleichsstudien verdeutlichen inzwischen die Notwendigkeit einer weitergehenden Lösung. So wurde im Pilotprojekt „Kleinräumiger Städtevergleich 2010“ (DifU 2012; Dohnke et al. 2012) versucht, auf kleinräumiger Ebene eine städtevergleichende Sozialraumanalyse zu rechnen. Das Projekt von Hartmuth Häußermann („res urbana“), dem Deutschen

Abbildung 3
Verteilung der Stadtteile des KOSTAT-Katalogs nach Einwohnerzahl, 2011

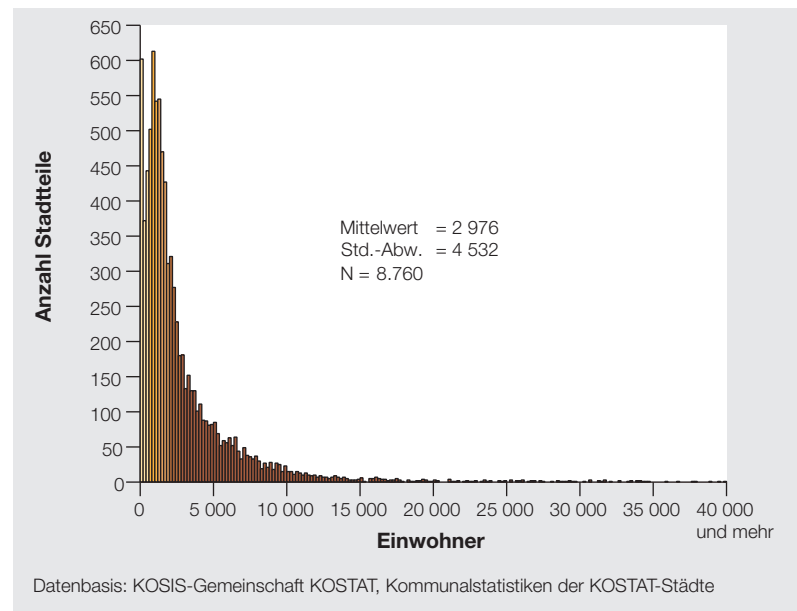
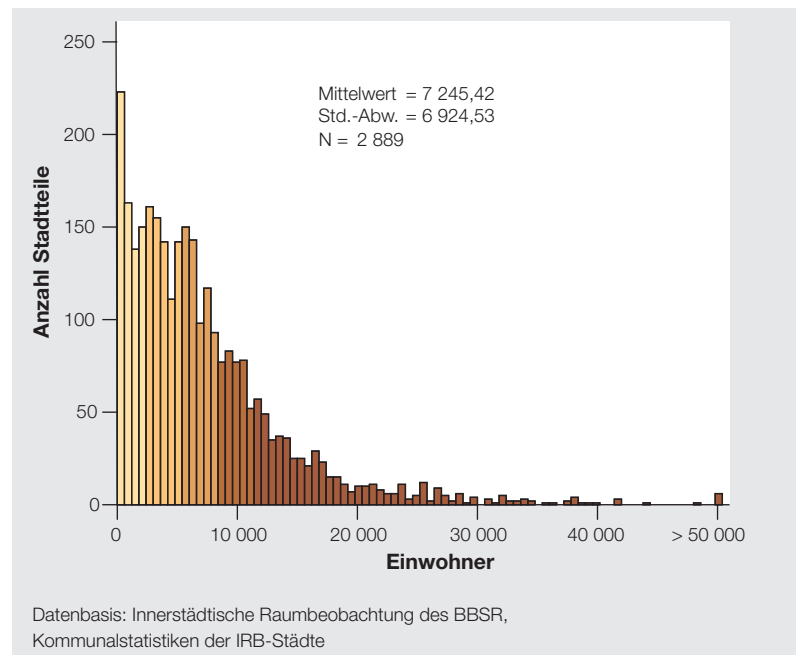


Abbildung 4
Verteilung der Stadtteile des IRB-Katalogs nach Einwohnerzahl, 2011



Institut für Urbanistik und 29 deutschen Städten stützte sich dabei auch auf die IRB. Angestrebt wurde eine Datenbasis für möglichst vergleichbare Teilräume. Um dies zu erreichen, wurden für einen Teil der Städte vorhandene Bezirke virtuell geteilt – unter der Annahme, dass die jeweils neu kreierten Teilbezirke identische Werte aufweisen. Für eine langfristige Beobachtung ist dieses Verfahren jedoch weniger geeignet.

Während wissenschaftliche Analysen bei Städtevergleichen auf Gesamtstadtebene sehr genau auf Größe und Struktur der Städte achten (können), ist diese notwendige Unterscheidung auf kleinräumiger Ebene nur durch Hilfskonstruktionen möglich. Insofern ist eine wesentliche Anforderung an eine kleinräumige, vergleichende Stadtforschung nur schwer umzusetzen. Derzeit sind zwei alternative Entwicklungen denkbar:

Zum einen könnte in den Städten geprüft werden, ob nicht eine Auswahl von Merkmalen standardmäßig nach Wahlbezirken bereitgestellt wird. Diese würden aufgrund ihrer mehr oder weniger vergleichbar großen Einwohnerzahl den durch die Stadtforschung geforderten homogenen Nachbarschaftsgebieten entsprechen. Problematisch ist dann allerdings die Verknüpfung mit den bisherigen Beobachtungseinheiten, da Wahlbezirke und statistische Raumeinheiten keine gemeinsamen Grenzen haben müssen. Weiterhin werden Wahlbezirke bei Änderung der Bevölkerungszahl regelmäßig neu zugeschnitten, so dass Zeitreihenanalysen extrem erschwert würden.

Zum anderen richten sich derzeit viele Hoffnungen auf eine zukünftig parallele Datenerhaltung – einerseits für die administrativen Raumeinheiten und andererseits für geografische Raster. Als statische Bezugseinheit haben Raster für Zeitvergleiche den Vorteil gleichbleibender Bezugsflächen. Weiterhin bieten sie die Möglichkeit der Verknüpfung mit diversen GIS-typischen Analyseverfahren. Als Nachteil wird derzeit diskutiert, dass sie für manche Fragestellungen zu starr sind und nicht nach nachbarschaftlichen Beziehungen, Lebenswelten bzw. räumlichen Gegebenheiten abgrenzen.

5.3 Auswahl, Definitionen und Klassenbildung der Merkmale

Beim Vergleich der Merkmalsprogramme der hier vorgestellten Datensammlungen fallen die unterschiedliche Auswahl der Merkmale, die im europäischen Vergleich teilweise verschiedenen Definitionen und die differenzierte Klassenbildung auf.

Die unterschiedliche Auswahl liegt vor allem im Zweck der Datensammlung begründet. So entstand die Innerstädtische

Raumbeobachtung als Kooperationsprojekt zum Aufbau eines Informationssystems auf kleinräumiger Ebene für Forschung und Politikberatung. Inhaltliche Fragestellungen der je aktuellen Stadtforschung standen und stehen damit im Vordergrund. So liegt der inhaltliche Schwerpunkt des Katalogs bei soziodemografischen und sozioökonomischen Merkmalen, die im Laufe des Bestehens verändert bzw. angepasst wurden. Das Merkmalspektrum ist sehr breit gefächert. Es beinhaltet nicht nur detaillierte Daten aus dem Einwohnermelderegister, sondern nutzt auch Daten der kommunalen Wohnungsregister und der Bundesagentur für Arbeit. Diese Vielfalt von insgesamt 400 Variablen können kleinere Kommunen nicht bereitstellen, da es ihnen an finanziellen, technischen wie personellen Ressourcen mangelt, um die entsprechenden Quellen kleinräumig aufzubereiten. Ohne abgeschottete Statistikstelle besitzen Kommunen auch keinen Zugang zu den entsprechenden Datenquellen. Insofern ist die IRB vor allem ein Großstadt-Katalog.

Anders sieht es für die AG KOSTAT aus. Ziel war es hier, kommunalstatistische Daten und kleinräumige Gliederungssysteme an unterschiedliche Zielgruppen zu verkaufen und sich damit auf dem Geodaten-Markt zu platzieren. Um möglichst viele Städte zu beteiligen, wurde die Merkmalsauswahl auf wenige nachgefragte Daten konzentriert. Der Datensatz wurde auf elf Variablen begrenzt und beinhaltet nur demografische Merkmale, die mittels Standardauswertungen aus dem Einwohnermeldewesen gewonnen werden können.

Die Datenauswahl von Urban Audit wird von der Generaldirektion Regionalpolitik bestimmt. Mit Hilfe der ausgewählten Indikatoren versucht sie, die aktuelle Situation in EU-Städten unter dem Aspekt der Lissabon-Strategie zu beurteilen. Daher werden Indikatoren gewählt, die die vorrangigen Ziele dieser Strategie – wie wissensbasiertes, ökologisches und wirtschaftliches Wachstum – abbilden können. Diese Ziele spiegeln sich aber hauptsächlich in den Merkmalen auf Ebene der Gesamtstadtdaten. Auf der kleinräumigen innerstädtischen Ebene werden aus Ressourcengründen nur wenige soziodemografische Merkmale angefordert. Wenn wir uns hier nur auf die kleinräumige Ebene der Sub-City-Districts beschränken,

liegt das Datenspektrum mit zukünftig nur noch 46 Variablen größtmäßig zwischen KOSTAT und IRB. Um vergleichbar zu sein, sind die nationalstatistischen Ämter verpflichtet, die Merkmale entsprechend den Definitionen des Statistischen Amtes der Europäischen Union (EuroStat) zu liefern. Die nationalen Daten müssen entsprechend angepasst oder umgeschätzt werden. Dadurch kann die Stadtforschung zwar erstmals auf vergleichbare Daten europäischer Städte zuzugreifen. Jedoch dürfen die UA-Daten nicht mit Daten gemäß nationaler Definition vermengt werden. Deshalb ist zwischen den hier vorgestellten Datensammlungen die Vergleichbarkeit nur zwischen IRB und Kostat gegeben.

6 Ausblick

Seit einiger Zeit wird es immer schwerer, die drei Datensammlungen aufrecht zu erhalten, da die Ressourcen für zusätzliche Aufgaben im Öffentlichen Dienst und insbesondere bei den Städten nur noch eingeschränkt zur Verfügung stehen. Der Einsatz wäre aber gerade auf Ebene der Kommunen von Bedeutung, da der aus den Datensammlungen zu ziehende Mehrwert den Arbeitsaufwand bei Weitem rechtfertigt. Alleine die Möglichkeit eines Vergleichs zwischen den Städten ist ein unschätzbares und notwendiges Hilfsmittel, um seitens der Kommune die Lage der eigenen Stadt einschätzen zu können und seitens des Bundes ein Verständnis für innerstädtische Prozesse zu erlangen.

Um die vorhandenen Ressourcen effektiver einsetzen zu können, gibt es aktuell seitens der KOSIS-Mitglieder eine Initiative, Datenformate und -inhalte sowie Liefertermine sowohl innerhalb der kommunalen Statistikstellen als auch bei den Datensammelnden Stellen stärker zu standardisieren. Als Zielvorstellungen sind im Gespräch:

- Entwicklung von Standard-Routinen zur Unterstützung der liefernden Städte;
- Konzeption eines Standard-Daten-Pakets mit einem Forschungsdatensatz;
- Standardisierung beziehungsweise Angleichungen kleinräumiger Gliederungen;
- Aufbau einer zentralen, Datensammelnden Stelle.

Auf jeden Fall sollen die angestrebten Neuerungen zu Arbeitserleichterungen in den kommunalen Statistikstellen führen, unter anderem, um mehr Zeit für gemeinsame Forschungsfragen und Analysen zu finden.